IN VITRO SELECTION OF DNAZYMES FOR COLORECTAL CANCER DETECTION

IN VITRO SELECTION OF FLUOROGENIC RNA-CLEAVING DNAZYMES FOR COLORECTAL CANCER DETECTION

BY

QIAN FENG

B. Sc. McMaster University, Hamilton, Ontario, 2012

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree

Master of Science

MASTER OF SCIENCE (2016) McMaster University

(Chemical Biology) Hamilton, Ontario

TITLE: In Vitro Selection of Fluorogenic RNA-cleaving DNAzymes for Colorectal Cancer Detection

AUTHOR: Qian Feng, B. Sc. (McMaster University)

SUPERVISOR: Dr. Yingfu Li

NUMBER OF PAGES: LXXIII, 73

Abstract

Colorectal cancer (CRC) is one of the leading causes of cancer death worldwide, affecting over one million individuals and causing 600,000 deaths annually. The mortality rate of CRC can be significantly reduced if it is detected early, suggesting the importance of cancer screening in CRC disease management. Currently, colonoscopy is the gold standard for CRC diagnosis as it is accurate and reliable. However, it is an invasive procedure that is associated with risks of complications, which contributes to the lack of patient compliance in colonoscopy screening. Other noninvasive detection methods suffer from poor sensitivity and specificity. Thus, there remains a great demand for the development of a non-invasive but cost-effective and accurate test for CRC diagnosis.

Recently, studies using next-generation sequencing techniques have revealed compositional changes in the intestinal microbiome associated with CRC, implicating the possibility of using fecal microbiome as potential diagnostic markers. Specifically, the level of the gram-negative bacterium, *Fusobacterium nucleatum,* has been shown to be elevated in CRC patients compared to healthy controls. The work described in this thesis aims to develop unique RNA-cleaving DNAzymes that can distinguish these molecular differences between healthy and CRC stool microbiomes.

RNA-cleaving DNAzymes are single stranded DNA molecules that have been isolated through in vitro selection to catalyze the cleavage of RNA. They are extensively used as analytical tools for metal ion sensing and bacterial detection. In our initial attempt of isolating RNA-cleaving DNAzymes directly against CRC patient fecal samples, we observed a significant level of non-specific nuclease degradation present in fecal samples

that interfered with the in vitro selection process. Alternatively, we conducted an in vitro selection experiment using the crude extracellular matrix produced by *F. nucleatum* and isolated an RNA-cleaving DNAzyme sensor, named RFD-FN1, that is activated by a protein marker by this bacterium. Importantly, the protein target of RFD-FN1 is resistant to heat denaturation where 50% protein remains functional after being heated at 90ºC for 20 min. RFD-FN1 is highly specific for *F. nucleatum* and it has a limit of detection of 10^7 CFU/mL without culture and a single cell when cultured for 36 hr. Even though the utility of RFD-FN1 in the presence of human fecal sample is still under optimization, the discovery of this novel molecular probe for *F. nucleatum* presents an important step forward towards the development of a novel DNAzyme-based detection method for colorectal cancer.

Acknowledgements

This project would not be possible without the support from my supervisor, Dr. Yingfu Li, for guiding me throughout the toughest times during the project. Thank you for sharing your knowledge and expertise and teaching me invaluable lessons on how to be a good scientist. You have provided me with a supportive and fun learning environment to grow professionally and personally.

I would like to thank my committee members Dr. Fred Capretta and Dr. Bruno Salena for making me think, pushing my limits and providing support and perspectives during my committee meetings. I would also like to thank Dr. Bruno Salena for providing the clinical samples for this project and the Surette lab for providing the anaerobic chamber and especially Jennifer Lau for providing the bacterial strains.

I would also like to thank Dr. Weijia Zhu for collaborating with me on the project; Dr. Monsur Ali for sharing your knowledge and experience in selection (my selection would not have worked without your help); Pushpinder Kanda for being my first mentor in the lab; Dr. Kha Tram and Dingran Chang for always being there for me, helping me troubleshoot and supporting me throughout the toughest times both professionally and personally. For all my friends at Li Lab, thank you for providing such a supportive and fun place for me to call home. Our times in the break room, during lab retreat and all the other activities are what kept me sane during the last two and half years.

Last but not least, I would like to thank my parents and all my friends for your unconditional love and always being there for me. I dedicate this thesis to you.

Table of Contents

Chapter 1. Introduction

Chapter 2. Materials and Methods

Chapter 3. Results and Discussions

List of Figures

List of Tables

List of Abbreviations

ATP	adenosine 5´-triphosphate
CEM	crude extracellular mixture
CFU	colony forming unit
CIM	crude intracellular mixture
CRC	colorectal cancer
DABCYL	4-(4-dimethylaminophenylazo) benzoic acid
ddH_2O	double-distilled water
DNA	deoxyribonucleic acid
DNAzyme	deoxyribonucleic acid enzyme
FOBT	fecal occult blood test
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
kD	kilo-Daltons
MES	2-(N-Morpholino)ethanesulfonic acid
NGS	next generation sequencing
PAGE	polyacrylamid gel electrophoresis
PCR	polymerase chain reaction
RFD	fluorogenic RNA-cleaving DNAzyme
RNA	ribonucleic acid
SELEX	systematic evolution of ligands by exponential enrichment
spp.	species

Chapter 1. Introduction

1.1 Colorectal Cancer

Colorectal cancer (CRC) is one of the leading causes of cancer death worldwide, accounting for over one million cases and 600,000 deaths annually.[1] Traditionally, colorectal cancer is considered a genetic disease where it is widely accepted that most sporadic CRC occurs through the adenoma-carcinoma sequence as proposed by Fearon and Vogelstein.[2] This model postulates that accumulation of genetic and epigenetic mutations drives the stepwise transformation of healthy mucosa towards pre-malignant lesions, with progression towards adenoma and eventually carcinoma over the course of many years. While the genetic and molecular mechanisms of CRC carcinogenesis have been extensively studied in the past, we know relatively little about its environmental causes.[3]

Recently, accumulating evidence has emerged to support the role of infectious agents as an important environmental factor in colorectal cancer. This is not surprising because our gastrointestinal tract is home to the majority of the microbes found in our body. It contains roughly 10^{14} microbial cells that make up more than 1000 different species.[4] These bacteria, collectively known as the intestinal microbiota, play key roles in orchestrating reactions important for host metabolism and immunity.[5] Early studies using mutant mouse models showed that mice that are kept under germ-free conditions develop significantly fewer tumours compared to when they have an intact microbiota.[6-8] Microbial dysbiosis has also been previously observed in CRC patients and specific

bacteria such as *Streptococcus bovis*, *Helicobacter pylori*, *Bacteroides fragilis* and *Escherichia coli* have been linked to CRC.[9-12]

In the last few years, studies using next-generation sequencing techniques have allowed us to identify changes in the bacterial composition associated with CRC, and revealed potential biomarkers for CRC detection. In the following chapters, I will present an overview of the recent next-generation sequencing data on human intestinal microbiota, emphasizing changes in bacterial composition during CRC progression. I will highlight research that support a role for the gut microbiota in CRC and discuss the utility of these findings in the development of a novel method for CRC detection.

1.2 CRC microbiome as potential diagnostics

In order to understand the impact of the intestinal microbiota on human health and its potential application in CRC detection, it is necessary to decipher the composition, diversity and the functions of this complex ecological system. In 2005, Eckburg and colleagues constructed the first large-scale map of the intestinal microbiota using sequence analysis of the bacterial small-subunit ribosomal RNA genes (16S rRNA genes) from colonic mucosal sites and feces of healthy individuals.[13] Since then, global efforts from the Human Microbiome Project and the Metagenomics of the Human Intestinal Tract Project were established to further characterize the constituents of our intestinal microbiota. The results reveal that there is a clearly distinguishable "core microbiome" that is present in all healthy individuals that share a similar set of functions.[4, 14] Deviations from this core microbiome could potentially lead to cancer. Because of the

diversity that exists among individuals, carriage of certain bacteria, as determined by the host genetics and lifestyle, may predispose the individual towards colorectal cancer. Therefore, identifying potential differences in the gut microbial community between healthy and CRC patients/tumour sites will help us elucidate the bacterial pathogenesis of CRC and reveal potential bacterial markers.

1.2.1 Mucosa-associated intestinal microbiome of CRC

From 2011 to 2012, three independent studies from Marchesi *et al.,* Kostic *et al.* and Castellarin *et al.* provided the first in-depth look of the colonic dysbiosis between CRC tumour tissues and adjacent healthy mucosa from four different patient cohorts.[15-17] Through bacterial 16S rRNA pyrosequencing, Marchesi *et al.* revealed that tumour tissues appeared to form a niche for members of the family Coriobacteriaceae and the genera *Roseburia, Fusobacterium* and *Faecalibacterium,* which are generally considered gut commensal bacteria. Interestingly, potential pathogenic bacteria such as members of the *Enterobacteriaceae*, specifically *Citrobacter, Shigella, Cronobacter, Kluyvera, Serratia* and *Salmonella* species were under-represented in tumours compared to adjacent healthy tissues.[15] Their results suggest that during CRC progression, the tumour microenvironment may be favourable for the colonization of gut commensals whereas pathogenic bacteria that may have initiated CRC are outcompeted at the advanced tumour stage. Using both 16S rRNA sequencing and metagenomic approaches, Kostic and colleagues observed an overall depletion of Firmicutes and Bacteoidetes while members of *Fusobacterium* species such as *F. nucleatum, F. mortiferum,* and *F. necrophorum* were

significantly enriched in cancerous tissues compared to adjacent off-tumour tissues.[16] Similarly, metatranscriptomic analysis from the Castellarin group further confirmed the enrichment of transcriptionally active *Fusobacterium* species in tumour as opposed to normal tissues.[17] Overall, these studies demonstrate that certain bacteria seem to preferentially inhabit either the tumour or adjacent healthy mucosa. Interestingly, all three studies showed an overabundance of *Fusobacterium* species in CRC tumour tissues. This is somewhat surprising because Fusobacteria are considered rare constituents of the intestinal microbiota and they have not been directly linked to CRC previously.

1.2.2 Luminal microbiome of CRC

From early studies by Eckburg and colleagues, we understand that the mucosal-adherent microbiota is distinct from the luminal (i.e. fecal) microbiota and they may fulfill different functions within the intestinal ecosystem.[13] Following the aforementioned studies on the mucosal-adherent microbiota, a number of studies have also focused on understanding the differences in the fecal microbiome between healthy subjects and CRC patients. In a study using 16S rRNA pyrosequencing techniques, Chen *et al.* compared both the mucosal-adherent microbiome, taken from rectal swab samples, and the luminal microbiome from stool samples between healthy and CRC patients.[18] Indeed, they found that the microbial composition of the intestinal lumen and mucosa differed significantly. In mucosal-associated microbiota, *Bifidobacterium, Faecalibacterium*, and *Blautia* were reduced while *Fusobacterium, Porphyromonas, Peptostreptococcus,* and *Mogibacterium* were enriched in CRC samples. With regards to

the microbiota of the intestinal lumen, members of the family Erysipelotrichaceae, Prevotellaceae, and Coriobacteriaceae were more abundant in cancer patients compared to healthy controls.[18] Because Erysipelotrichaceae, Prevotellaceae, and Coriobacteriaceae are all related to metabolic disorders (such as obesity) and energy metabolism, it is plausible that the luminal microbiota influences CRC risk through metabolic exchange with the host.[19,20] Moreover, Wang *et al.* showed a significant depletion of butyrate-producing bacteria such as *Roseburia* in the feces of CRC patients and it has been shown that butyrate has protective properties against CRC.[21-23] Altogether, these studies suggest that the dysbiosis of the luminal microbiota may be influencing CRC progression through metabolic pathways.

1.2.3 *Fusobacterium* as a candidate bacterial marker for CRC

Perhaps one of the most intriguing discoveries that emerged from all the microbiome sequencing studies is the prevalence of *Fusobacterium* species associated with colorectal cancer. *Fusobacterium* are Gram-negative, anaerobic bacteria that usually inhabit the oral cavity and they have been implicated in various periodontal diseases.[24-26] Research has shown that *Fusobacterium* spp. are enriched in CRC tumours compared to normal colons, as well as in the feces of CRC patients compared to healthy controls.[15-18, 27, 28] These findings raise the question of whether *Fusobacterium* is a causative agent of CRC or an opportunistic colonizer. Studies on colorectal adenomas have found that *Fusobacterium* is also enriched in adenoma tissues and fecal samples of

CRC patients, suggesting a role for *Fusobacterium* in the early stages of CRC development.[26, 29, 30]

Previously, *Fusobacterium* has been shown to be highly invasive and adherent to mucosal surfaces.[25, 31] It is also associated with inflammatory bowl diseases, which is a known risk factor for CRC.[31] Therefore, it is possible that *Fusobacterium* contributes to CRC progression through inflammatory pathways. Using the multiple intestinal neoplasia mouse model, Kostic *et al.* demonstrated that *Fusobacterium nucleatum* increased tumour multiplicity, and that tumours from mice exposed to *F. nucleatum* exhibited a pro-inflammatory expression profile that was also present in human fusobacteria-positive tumours.[30] Moreover, Rubinstein *et al.* further demonstrated that *F. nucleatum* used its FadA adhesion to attach to E-cadherin on CRC cells, thereby activating downstream pro-oncogenic and inflammatory pathways.[24] These results suggest that *Fusobacterium* promotes CRC through mucosal inflammation. Due to the complexity of the intestinal environment and the possible contributions from other bacterial pathogens, more studies need to be conducted to confirm such an association. Nonetheless, because of its high prevalence in colorectal adenoma and carcinoma tissues/feces, *Fusobacterium* is an attractive candidate marker for CRC diagnosis.

1.2.4 Fecal microbiome as potential CRC diagnostics

The availability of next-generation sequencing techniques has allowed researchers to gain a comprehensive view of the microbial dysbiosis associated with CRC. Even though these results cannot confirm the causal relationship between gut microbes and tumourigenesis, they have enabled the possibility of using microbiota as a diagnostic method for CRC.

Recently, Zeller *et al.* demonstrated that by using metagenomic sequencing of fecal samples from 156 participants, they identified a set of taxonomic markers that distinguished CRC patients from healthy controls.[32] The accuracy of their metagenomic detection method was comparable to that of the fecal occult blood test (FOBT), which is a widely used noninvasive method for CRC detection. When combined with the FOBT, the metagenomic detection improved sensitivity by approximately 45% relative to the FOBT alone. However, this method was not able to distinguish between early- and late-stage CRC and therefore it has limited uses in early CRC diagnosis. Similarly, by combining the microbiome data with known risk factors of CRC (i.e. body mass index, age and race), Zackular and colleagues developed a model that improved the predictive power for both adenoma and carcinoma patients.[33]

Together, these studies demonstrate the feasibility of using stool microbiome composition as a diagnostic method for colorectal cancer. With further optimization and validation studies using a larger and more diverse group of individuals, this approach can potentially change the way colorectal cancer is screened.

1.3 Current detection of CRC

The detection of colorectal cancer is extremely crucial because CRC is a highly treatable if it is detected early. Its mortality rate changes drastically depending on the stage of cancer at the time of diagnosis. While patients with localized cancer (stage I) have a 90% 5-year survival rate, prognosis worsens significantly when the cancer metastasizes to nearby lymph nodes or distant organs, as the 5-year survival rate drops to 12%.[34] Such striking differences in survival between early and advanced disease suggests that early detection is the key to reduce CRC-related deaths. However, despite the availability of CRC screening programs in Canada, screening rates remain low, where it is estimated that less than 40% of CRC cases are detected before cancer metastasis. [35, 36]

Currently, colonoscopy is widely accepted as the gold standard method for CRC diagnosis. It can detect both cancer and precancerous lesions with high reliability, and it allows the removal of detected polyps at the same time. Results from case-control and cohort studies showed significant reductions of about 50% on CRC incidence through colonoscopy screening.[37] Therefore, it is recommended that individuals over the age of 50 should undergo colonoscopies every 10 years.[35] Despite the availability of such an effective diagnostic test, compliance to colonoscopy screening is poor. This is attributed to the invasive nature of the procedure, which involves intensive bowl cleaning, sedation and the risk of serious complications such as bowl perforation and bleeding.[38] Overcoming this barrier is important for the reduction of CRC mortality and incidence.

Among the noninvasive tests for CRC, guaiac fecal occult blood test (FOBT) is widely used as an early screening program in Canada. It is a colorimetric, paper-based test strip that detects blood in stool as a result of fragile blood vessels on polyps or cancerous tissues in CRC patients. However, the utility of this test is limited due to lack of sensitivity and high false-positive rate, which imposes additional economic burden on the healthcare system.[39] Similarly, other noninvasive tests such as immunochemical FOBT also suffer from poor sensitivity and specificity for CRC. The advantages and disadvantages of current CRC screening methods have been extensively reviewed in the past and they are summarized in Table 1.[40, 41] Altogether, there remains a great demand for the development of more cost-effective, accurate and noninvasive test for CRC diagnosis.

The purpose of my thesis is to develop a novel detection method using catalytic DNA molecules that can distinguish the molecular differences between healthy and CRC stool microbiomes. The rationale and experimental design of my study will be discussed in the following chapters.

Table 1. Comparison of the current CRC diagnostic methods.

Test	Marker	Advantages	Disadvantages
Invasive Tests			
Colonoscopy	Cancerous lesions & polyps in the entire colon	View of the entire colon and rectum; Biopsy and removal of polyps during the test; Most sensitive test available	Small polyps, nonpolypoid lesions hard to detect; Extensive bowl cleaning before test; Sedation; Risk of complications
Flexible Sigmoidoscopy	Cancerous lesions & polyps in sigmoid colon and rectum	Quicker and fewer complications than colonoscopy; Minimal discomfort; Less extensive bowl cleansing	Cannot detect polys in upper colon May require colonoscopy as follow-up; Risk of complications
CT Colonography (virtual colonoscopy)	Cancerous lesions & polyps in the entire colon using X-ray	View of the entire colon and rectum; No risk of complications	Exposure to X-ray; Small polyps, nonpolypoid lesions hard to detect; Extensive bowl cleaning before test
Double contrast barium enema (DCBE)	Cancerous lesions & polyps in the entire colon using X-ray	View of the entire colon and rectum; Low risk of complications; No sedation required	Exposure to radiation; Small polyps, nonpolypoid lesions hard to detect; Extensive bowl cleaning before test; Possible false-positives
Noninvasive Tests			
Guaiac Fecal Occult Blood Test (FOBT)	Heme	Inexpensive; Proven to decrease CRC mortality;	Poor sensitivity; Not specific for CRC (high false-positives); Require dietary & medication restrictions;
Immunochemical FOBT	Globin	Inexpensive; No dietary & medication restrictions	Poor sensitivity; Not specific for CRC (high false-positives);
Multi-target DNA in stool	Mutations in K-ras, p53, APC, BAT-26 and long DNA	Improved sensitivity to FOBT	Much higher cost than FOBT

1.4. Probing CRC fecal microbiome using catalytic DNA

1.4.1 Catalytic DNA

It is now widely accepted that single-stranded DNA molecules can adopt higher degree structures to perform molecular functions including target recognition and enzymatic catalysis.[42, 43] Those that can bind to specific targets are named DNA aptamers whereas those that can increase the rate of chemical transformations are termed DNAzymes. These functional DNA molecules are not found in nature, but rather developed in the laboratory through a process called in vitro selection, also known as Systematic Evolution of Ligands by Exponential Enrichment (SELEX).[44-46] In vitro selection is a combinatorial technique that recapitulates Darwinian evolution in a test tube where a specific selective pressure is exerted on a pool of heterogeneous nucleic acid molecules. Through iterative rounds of selective amplification based on physical or biochemical properties, only the species with desired traits will be enriched in the population, as shown in Figure 1. The result of this “test-tube” evolution is the generation of DNA molecules with a variety of fascinating functions.

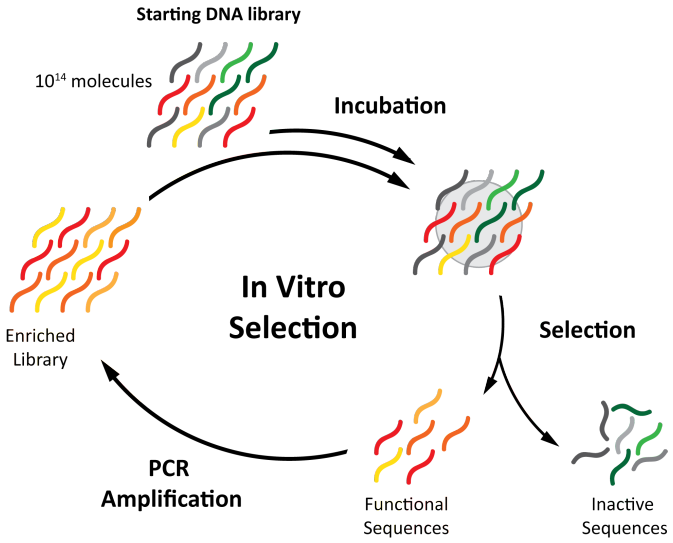

Figure 1. *General in vitro selection scheme*. A starting library of ~ 10^{14} unique sequences is subjected to incubation under a certain selective pressure. The inactive sequences are removed while functional sequences are regenerated through PCR amplification. This process is repeated many times until the nucleic acid pool is dominated with functional sequences.

Over the last 25 years, a vast amount of research on functional DNA molecules have highlighted the utility of DNA aptamers and DNAzymes as robust molecular tools, owing to their precise target recognition and robust enzymatic capabilities. Specifically, DNA aptamers and DNAzymes are highly attractive tools in biosensing applications.[47-49] This is because DNA synthesis can be achieved through automated solid-phase synthesis at a low cost, and it eliminates biological contaminants such as viruses and bacteria compared to antibody-based biosensing platforms. Additionally, DNA molecules are chemically more stable than their protein counterparts, allowing for longer shelf lives and easy handling. Therefore, our lab has been interested in using in vitro selection to derive DNA aptamers and DNAzymes for biosensing purposes.

1.4.2 In vitro selection of catalytic DNA molecules

1.4.2.1 In vitro selection

Ever since its conception in the early 1990s, in vitro selection has been a powerful tool for generating many different molecular probes. Its unique advantages in biosensing applications can be attributed to the following traits. In general, in vitro selection starts with a pool of chemically synthesized sequences that is completely random. This allows for the maximal sampling of the sequence space, so that a variety of different sequences and adaptation routes can be explored under a given selection pressure. In a single experiment, one can sample around 10^{13} - 10^{16} sequences simultaneously, which increases the probability of finding functional molecules. In addition, almost all experimental parameters throughout the selection process such as incubation time, selection stringency, target specificity and reaction rate can be tailored towards an experimenter's specific needs.[50, 51] The flexibility in the experimental design allows researchers to customize its selection scheme for different downstream biosensing applications.

Both DNA aptamers and DNAzymes have been equally explored as suitable molecular probes for biosensor developments. The choice of DNA aptamers versus DNAzymes for a specific application often depends on personal preference and the nature of downstream assay development. For example, DNA aptamers have the unique feature of recognizing a wide range of analytes such as small molecules, proteins, and even whole cells.[47, 51] However, a recognition molecule is only one part of a biosensor as the DNA aptamer platform lacks a transduction element, which is required for signal generation. Conversely, DNAzymes can be an advantageous alternative because signal

generation can be incorporated into the DNAzyme. This is reported in literature for fluorescent RNA-cleaving DNAzymes of metal ions.[52] Since DNAzymes require metal ions as cofactors for catalysis, in vitro selection experiments can be designed to isolate highly specific DNAzymes that is dependent on the metal ion of interest.[47] However, when isolating DNAzymes that are responsive to a specific target other than metal ions, the selection process is much harder because the DNA molecule must perform both target recognition and enzymatic catalysis.

1.4.2.2 In vitro selection of RNA-cleaving DNAzymes

Among the plethora of chemical transformations that have been developed to date for DNAzymes, RNA-cleavage is the most broadly used chemistry for biosensing applications. RNA-cleavage refers to the transesterification reaction of RNA where a 2′-hydroxyl group of ribose initiates a nucleophilic attack to the neighboring phosphodiester bond. This results in the formation of a 2′, 3′-cyclic phosphate and a free 5′-hydroxyl terminus, as shown in Figure 2.[53] This chemical reaction has been extensively studied in the past due to the availability of many naturally occurring RNA-cleaving ribozymes (RNA counterparts to DNAzymes), as well as many classes of man-made RNA-cleaving DNAzymes.[47, 52, 54-58] Consequently, the wealth of knowledge on the structural and catalytic properties of these DNAzymes significantly facilitates the biosensor development process. RNA-cleaving DNAzymes are also the most catalytically robust enzymes among all DNA enzymes.[48, 59, 60] This is highly favourable in biosensors because a faster catalyst is directly translated into a quicker response time. Moreover,

RNA cleavage results in the separation of two shorter nucleic acid strands, which can be easily coupled to various signal transduction platforms to generate electrochemical, colourimetric or fluorescent signals.

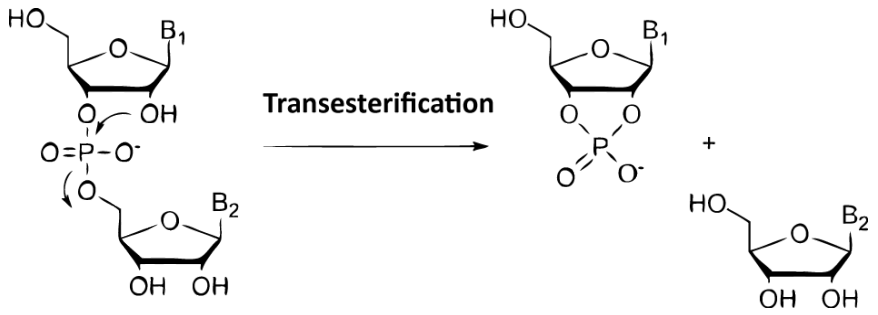

Figure 2. *Reaction mechanism of RNA cleavage.*

Due to nucleic acids' intrinsic affinity for metal ions, in vitro selection of RNA-cleaving DNAzymes often requires the presence of metal ions as cofactors in the transesterification reaction. In fact, the first DNAzyme isolated was a Pb^{2+} dependent RNA-cleaving DNAzyme that cleaves a single ribonucleotide in an all DNA sequence.[61] Since then, many classes of RNA-cleaving DNAzymes have been engineered as biosensors for different metal ions.[47] These DNAzymes has also been expanded to cleave all RNA substrates such as the 10-23 and 8-17 DNAzymes.[62] Most recently, there is increasing interest in extending the application of RNA-cleaving DNAzymes to detect more complex targets such as small molecule metabolites, and even bacterial and mammalian targets.[47, 63-66] Notably, our lab has been a pioneer in developing RNA-cleaving DNAzymes for bacterial detection.

1.4.2.3 In vitro selection of RNA-cleaving DNAzymes for bacterial detection

Our lab has been interested in developing a special class of RNA-cleaving DNAzymes that couples RNA cleavage and fluorescent signaling in a single step. This is achieved through the design of a chimeric DNA/RNA substrate where a single ribonucleotide is flanked by two nucleotides modified with a fluorophore and a quencher as shown in Figure 3.[52] In the presence of the intended target, an RNA-cleaving DNAzyme will induce the cleavage at the ribonucleotide junction, which releases the fluorophore-labeled DNA strand from the quencher strand. As s result, target recognition induces RNA cleavage, which is directly linked to fluorescent signal generation in a "mix-and-read" type assay (Figure 3). This class of DNAzymes is named fluorogenic RNA-cleaving DNAzymes or RFDs. Since 2003, our lab has successfully selected many RFDs with varying catalytic and signaling properties.[52, 67-70] More recently, we began to expand the utilities of RFDs in detecting complex targets such as bacterial pathogens.

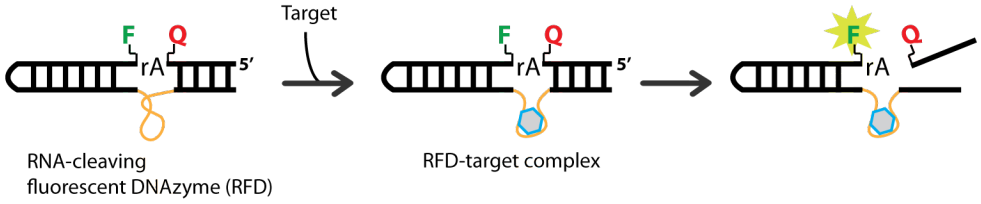

Figure 3. *Schematic of fluorogenic RNA-cleaving DNAzyme*. The DNAzyme component of the RFD is shown in orange and is ligated to the substrate at the 5′ end. The substrate contains a single adenosine ribonucleotide (rA) flanked by a pair of deoxyribothymidine modified with fluorescein (F) and DABCYL (Q). In the presence of the target, the DNAzyme catalyzes the cleavage at the ribonucleotide junction and allows the departure of the quencher from the fluorophore. This induces an increase in fluorescent intensity that can be measured at 521 nm.

In conventional biosensor development, it is important to first establish a set of biomarkers that can signal the presence of a particular disease or target. However, the biomarker identification process is often the rate-limiting step because it is tedious, slow and expensive to conduct. To circumvent this problem, our group developed a novel in vitro selection strategy from which RFDs are directly isolated using unpurified cellular mixtures as the target.[71] The outcome of this method is the generation of RFDs that bypasses the biomarker discovery step while still maintaining the ability for highly selective detection. There are two important concepts behind this approach. First, it is a "many-against-many" method where the randomized DNA library provides many potential RFDs while the unpurified cellular matrix provides many potential biomarkers (Figure 4). Second, because we are bypassing biomarker discovery, which insures target specificity, it is crucial to incorporate stringent counter selection to eliminate any nonspecific sequences. Together, highly specific RFDs can be developed for a wide variety of biological targets using this in vitro selection strategy.

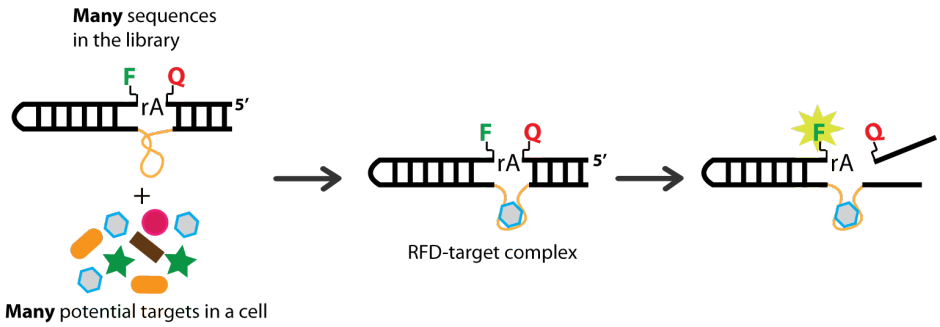

Figure 4. *Schematic of the "many-against-many" approach for deriving RFDs for biological targets.* The DNA library provides many different DNA sequences whereas a cell produces many potential targets such as small molecules and proteins.

The first example of this strategy was the development of a specific RFD capable of recognizing *Escherichia coli*.[63] In this study, the crude extracellular mixture (CEM) left behind in the *E. coli* culture was used as the positive selection target. Living microbes, grown under nutritious conditions, constantly exchange materials with their environments and leave behind a mixture of small and macromolecular molecules that can be highly characteristic for each bacterium. Using CEM from *E. coli* (CEM-EC) as the positive selection target and CEM from *Bacillus subtilis* (CEM-BS) as the counter selection target, an RFD, named RFD-EC1, was isolated to respond specifically to CEM-EC and no other bacterial CEMs. RFD-EC1 also demonstrated a strong detection sensitivity of as low as 100 *E. coli* cells.

In another study, the concept of crude cellular mixture was applied to mammalian cells where the crude cell lysate of a triple-negative breast cancer cell line MDA-MB-231 was used as the selection target.[65] Using the same strategy of positive selection and counter selection, an RFD named AAI2-5 was isolated to specifically detect MDA-MB-231 cell line at a detection sensitivity equivalent of 5000 cells/mL.

Most recently, a highly strain-specific RFD named RFD-CD1 that is only activated in the presence of *Clostridium difficile* strain BI/027 CEM and not any other *C. difficile* strains was isolated.[66] This was achieved through positive selection using CEM from *C. difficile* BI/027 and counter selection using CEMs from *E. coli*, *B. subtilis* and *C. difficile* CD630 (a non-BI/027 strain).

Together, all three studies demonstrate the feasibility of our novel "many-against-many" in vitro selection strategy. This method can be generalized to develop RFDs for

many different bacterial pathogens and cell lines. By adjusting the selective pressure, we can fine-tune these RFDs to respond to a specific strain of bacteria or a general category of bacteria depending on downstream applications.

In this thesis, I will describe the use of our unique in vitro selection strategy to isolate a specific RFD for *Fusobacterium nucleatum* detection. As mentioned in Chapter 1.2.4, *F. nucleatum* has emerged in recent years as a highly attractive marker for colorectal cancer detection. However, current methods for *F. nucleatum* detection are inadequate for providing simple and fast diagnosis. The gold standard culture-based detection requires a cultivation step of up to 96 hr and many time-consuming biochemical tests.[72-74] Polymerase chain reaction (PCR) and real-time PCR provides shorter detection time but still require expensive equipment and trained technicians.[73-75] Therefore, innovative technologies using RFDs will provide inexpensive, specific and easy-to-use detection for *F. nucleatum* as a potential application for colorectal cancer diagnosis.

1.4.2.4 In vitro selection of RNA-cleaving DNAzymes for complex biological samples

The success of our RFDs for bacterial and mammalian cell detection has prompted us to develop RFDs for even more complicated targets such as human stool samples. Human stool is an attractive detection target because it can be easily collected as a noninvasive procedure. From the research studies presented in this thesis, it is evident that there are distinguishable differences between the fecal microbiota of CRC patients and that of healthy controls. We hypothesize that certain proteins or metabolites produced by

these bacteria or by the host in response to colonic colonization by these bacteria can serve as molecular signatures of CRC and that these signatures can be found in the fecal samples of CRC patients. Therefore, we propose to develop RFDs through in vitro selection where stools samples from clinically confirmed CRC patients are used as positive selection targets and those from healthy controls are used as counter selection targets. The success of this project will have three important implications: 1) the development of a noninvasive, accurate and cost-effective alternative for CRC detection, 2) the use of complex human biological samples directly as a target in selection is unprecedented and it extends the applications of *in vitro* selection, and 3) the possibility of identifying novel CRC biomarkers through our RFD biosensor, which will provide insights into CRC diagnosis and treatment.

1.5 Thesis Objectives

The objective of my thesis is to apply in vitro selection to develop novel fluorogenic RNA-cleaving DNAzymes for colorectal cancer detection. To achieve this goal, we have designed a two-pronged approach: 1) to isolate CRC-specific RFDs through in vitro selection using CRC patient fecal samples as target; 2) to isolate *Fusobacterium nucleatum*-specific RFDs using *F. nucleatum* CEM as the selection target. Both strategies present a new platform for developing simple, accurate and cost-effective detection for colorectal cancer.

Chapter 2. Materials and Methods

2.1 Materials and Methods for CRC Selections

2.1.1 DNA oligonucleotides synthesis and purification

DNA oligonucleotides were purchased from IDT DNA Technologies (Coralville, USA) with the exception of FQ30, which was purchased from Keck Oligo Synthesis Facility (Yale University). All oligonucleotides were synthesized through automated DNA synthesis using standard phosphoramidite chemistry and purified by 10% denaturing polyacrylamide gel electrophoresis (dPAGE). DNA concentration was measured at 260 nm using any UV-vis spectrophotometer, such as NanoVue (GE Healthcare, UK). The sequences of all oligonucleotides are listed in Table 2.

2.1.2 Enzymes and chemical reagents

Radioactive nucleotides were purchased from Perkin Elmer (Massachusetts, USA). Enzymes were purchased from MBI Fermentas (Burlington, ON, Canada) with the exception of DNA polymerase, which is purchased from Biotools (Biotools, Spain). Proteinase K was purchased from Thermo Scientific (Massachusetts, USA). Non-radioactive nucleotides are purchased from Fermentas. All other chemicals are purchased from Sigma-Aldrich (Missouri, USA) unless specified otherwise. All multi-component solutions are made in-house using autoclaved, deionized water, filtered through the Milli-Q® Intergral Water Purification System (EMD Millipore, Massachusetts, USA).

Table 2. Sequences used for two CRC selections.

Name	Sequence
Fluorogenic Susbtrate (FQ30)	5′-CTATGAACTGACF-rA-QGACCTCACTACCAAG (F=fluorescein-labeled dT, Q=DABCYL-labeled dT, rA= adenine ribonucleotide)
Selection-CRC1	
Library (L1)	5′- GAACAGAGCAGACTAACG-N_{50}-CGAGTCAGTCAGTAAG TC (N= 25% A, 25% G, 25% C, 25%T)
Forward Primer (L1-FP)	5′- GAACAGAGCAGACTAACG
Reverse Primer 1 (L1-RP1)	5′- GACTTACTGACTGACTCG
Reverse Primer 2 (L1-RP2)	5′- TTTTTTTTTTTTTTTTTTTT-S_{18}-GACTTACTGACTGACTC G (S_{18} is a 18-atom hexa-ethyleneglycol spacer)
Ligation Template (L1-T1)	5′- TCTGCTCTGTTCCTTGGTAGTGAG
Deep Sequencing Forward Primer (L1-SFP)	5′- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT ACACGACGCTCTTCCGATCTGAACAGAGCAGACTAACG
Deep sequencing Reverse Primer (L1-SRP)	5′- CAAGCAGAAGACGGCATACGAGATATTCTCGTGACTGG AGTTCAGACGTGTGCTCTTCCGATCTGACTTACTGACTGAC TCG
Selection-CRC2	
Library (L2)	5′- AACGCTTACAATGACACTCC-N_{50}-GCGTGCGTAATAGTG TCAAG (N= 25% A, 25% G, 25% C, 25%T)
Forward Primer (L2-FP)	5′- AACGCTTACAATGACACTCC
Reverse Primer 1 (L2-RP1)	5′- CTTGACACTATTACGCACGC
Reverse Primer 2 (L2-RP2)	5′- TTTTTTTTTTTTTTTTTTTT-S_{18}-CTTGACACTATTACGCAC GC (S_{18} is a 18-atom hexa-ethyleneglycol spacer)
Ligation Template (L2-T1)	5′- ATTGTAAGCGTTCTTGGTAGTGAG
Deep Sequencing Forward Primer (L2-SFP)	5′- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT ACACGACGCTCTTCCGATCTAACGCTTACAATGACACTCC
Deep sequencing Reverse Primer (L2-SRP)	5′- CAAGCAGAAGACGGCATACGAGATATTCTCGTGACTGG AGTTCAGACGTGTGCTCTTCCGATCTCTTGACACTATTACG CACGC

2.1.3 CRC fecal sample preparation

With the assistance of Dr. Bruno Salena from McMaster Health Sciences, we collected fecal specimens from 15 clinically confirmed, treatment-naïve CRC patients. One patient specimen (named P2) was excluded from the study because the patient was diagnosed with a rare squamous cell carcinoma rather than adenocarcinoma, which is the most common CRC cell type. We also obtained fecal samples from 4 healthy control patients who do not present abnormalities in colonoscopies performed within 12 months before sample collection. All fecal samples were stored at -80°C before processing. During target extraction, about 5 grams of thawed fecal samples were added to 10 mL of 1× selection buffer (300 mM NaCl, 100 mM KCl, 10 mM $MgCl_2$, 5 mM $MnCl_2$, 60 mM HEPES, pH 7.5, 0.01% Tween 20) with cOmplete protease inhibitor cocktail (Roche). These mixtures were shaken at room temperature for 2 hr and spun down using VWR clinical-100 centrifuge at 5000 rpm for 15 min. The supernatants were sonicated for 30 min using Branson 2510 Ultrasonic Cleaner and further spun down using a tabletop centrifuge. In order to limit the possible target range to between 10 kilo-Daltons (kD) and a 100 kD, the supernatants were filtered through a 100 kD spin column, followed by a 10 kD spin column. Protein concentrations of the recovered targets were determined using Bradford Assay and normalized to around 1 mg/mL. The extracted samples were stored in -80°C with 20% glycerol until further use.

2.1.4 Overall selection scheme

A randomized 50-nucleotide DNA library, flanked by two primer binding arms, was used for in vitro selection of a CRC-specific RFD. The library was first radioactively labeled at the 5′ end and ligated to FQ30. Ligated library was subsequently purified by 10% polyacrylamide gel electrophoresis (PAGE) and incubated with solubilized CRC fecal targets to facilitate catalysis of RNA cleavage. Inactive sequences were removed by dPAGE based on the length difference between the cleaved and uncleaved sequences. Active sequences were amplified through polymerase chain reaction (PCR) and the sense strand obtained through a second PCR reaction. Following PAGE purification, an enriched DNA pool was obtained and the cycle was repeated until the desired RNA-cleaving DNAzyme was found. A detailed schematic is illustrated in Figure 2.

2.1.5 Selection-CRC1

2 nmol of the DNA library (L1) was first ligated to equimolar amounts of FQ30 using T4 DNA ligase. In the first round of selection, 30 µL of CRC mix A containing fecal extracts from patient 1, 3, 4, 5 & 6 was incubated with the ligated library for 4 hr at room temperature in 1× selection buffer at a final volume of 150 µL. All subsequent selection rounds differ from the first round only by the selection time and the target used, as shown in Table 2. After round 3 of positive selection, a counter-selection round was introduced to improve the RFD specificity for CRC fecal samples. The healthy patient sample mix was incubated with the DNA pool for 2 hr to eliminate any cleaved sequences. Refer to Table 2 for the detailed selection protocol.

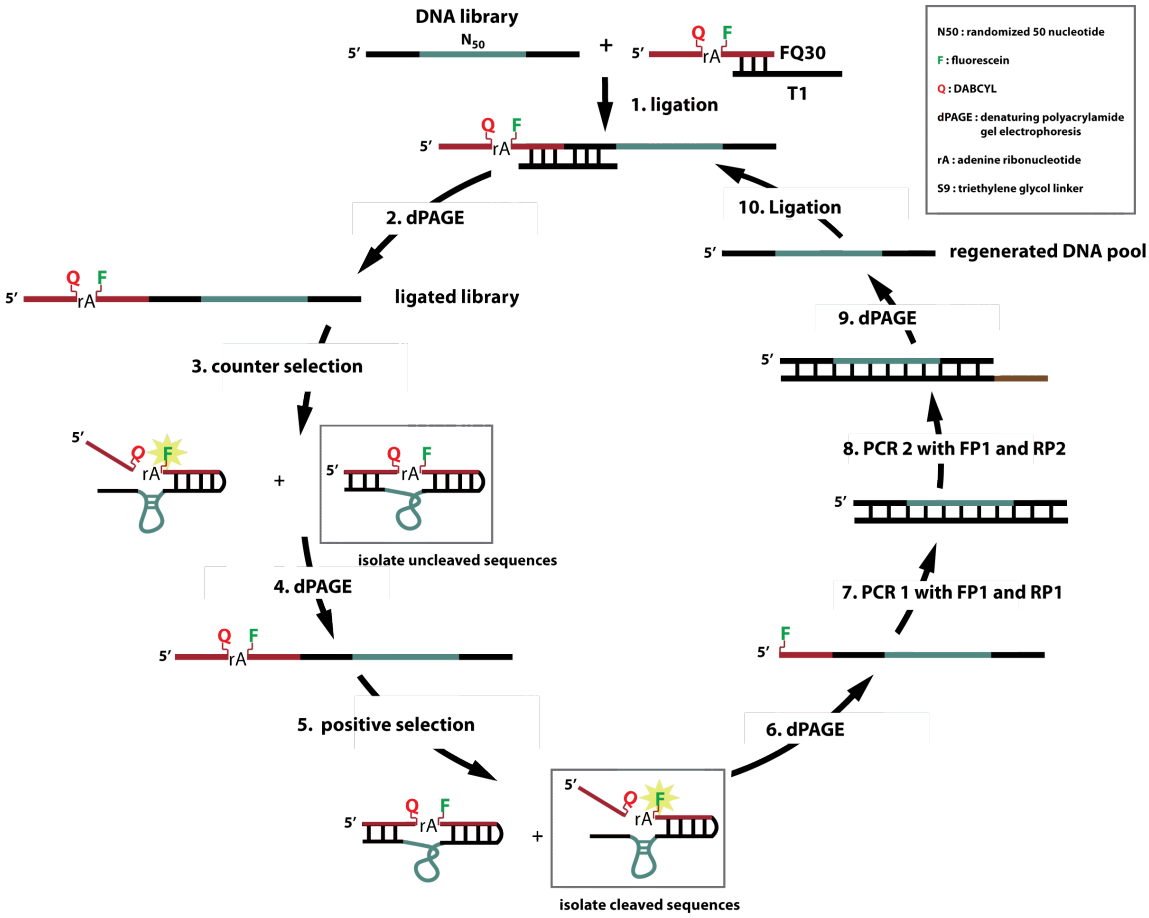

Figure 5. *Detailed in vitro selection scheme.* A randomized 50 nucleotides DNA library was first phosphorylated at the 5′ end and ligated to the substrate (1). The substrate contains a cleavage site at the adenosine ribonucleotide (rA), which is flanked by a fluorescein-labeled deoxyribothymidine on the 3′ and a DABCYL-labeled deoxyribothymidine on the 5′ end. The ligated products were purified by denaturing PAGE (2) and incubated in counter selection target to eliminate any unspecific DNAzymes (3). Uncleaved sequences from counter selection were purified (4) and incubated with positive selection target to isolate RNA-cleaving DNAzymes that are responsive to the target of interest (5). Cleaved sequences were then isolated by dPAGE (6) and amplified by PCR (7). Another PCR reaction was performed to regenerate the sense strand (8). Following dPAGE purification (9), an enriched DNA pool was obtained and the cycle was repeated until the desired RNA-cleaving DNAzymes were found.

Table 3. Selection-CRC1 protocol.

Target mix	Patient
CRC mix A	P1, P3, P4, P5, P6
CRC mix B	P7, P8, P9, P10, P11
CRC mix C	P12, P13, P14, P15
Healthy Mix	H1, H2, H3, H4

Round	Selection type	Target	Incubation time
1	positive	CRC mix A	4h
2		CRC mix A	2h
3		CRC mix B	1h
Counter-1	counter	Healthy mix	2h
4	positive	CRC mix B	1h
5		CRC mix C	1h
6		CRC mix C	1h
7		CRC mix A	1h

2.1.6 Nuclease degradation analysis

The class 1 sequence from round 4 sequencing results from the selection-CRC1 was used to analyze nuclease degradation. An 18-nucleotide polyT sequence containing three phosphorothioate linkages at the 5′ end was ligated to FQ30 as shown in Figure 5-A. It was further ligated to the class 1 sequence from round 4. The full-length phosphorothiolated sequence and the unmodified sequence were incubated with one CRC fecal sample (from patient 7) for 1 hr at room temperature in 1× selection buffer. The phosphorothioated sequence was also ligated to the DNA library used in the original selection. The modified library and the original library were tested with one CRC sample from P7 for up to 24 hr in order to evaluate the degree of nuclease degradation.

2.1.7 Selection-CRC2

2 nmol of a re-designed 50-nucleotide DNA library (L2) with a new set of primer binding arms was used in this selection to avoid cross-contamination with the previous selection. All fecal extracts from CRC patients were pooled into one group as the target for positive selection, while healthy patient fecal extracts were pooled together for counter-selection. In order to eliminate nuclease degradation, all targets were heated at 90°C for 10 min and cooled to room temperature before adding them to the selection reaction. An overall of seven rounds of positive selection were performed. For the first round of selection, 10 μL of pooled CRC patient extracts were incubated with the library for 4 hr at room temperature in 1× selection buffer at a final volume of 100 μL. From round 2 to round 7, 2 μL of CRC patient extracts were used at a final volume of 15 μL. The reaction times were 2 hr for round 2 and 1 hr for round 3 to 7.

2.1.8 Sequencing analysis

PCR products of round 4 & 7 from both selection-CRC1 and selection-CRC2 experiments were appended with additional 5′ extensions on both the forward and reverse primers via PCR. Primer set L1-SFP and L1-SRP was used for sequencing selection-CRC1 results and primer set L2-SFP and L2-SRP was used for sequencing selection-CRC2 results. PCR products were separated on agarose gels, purified, and sent for sequencing using the Illumina MiSeq sequencer. Sequence reads from MiSeq were first trimmed of their primers using Geneious. The resulting 50 nt reads were filtered for quality using PrinSeq v0.20.4 to make sure only high quality reads were used for further

analysis.[76] All sequences containing any bases with Phred scores < 20 (base-call probability < 99%) were removed. Using a clustering algorithm CD-HIT-EST, same sequences within each pool were grouped into classes and then ranked by the number of sequences in each class to identify the dominating sequences in the pool.[77]

2.2 Materials and Methods for *Fusobacterium nucleatum* Selection

2.2.1 DNA oligonucleotides synthesis and purification

DNA oligonucleotides were purchased from IDT DNA Technologies (Coralville, USA) with the exception of FQ30, which was purchased from Keck Oligo Synthesis Facility (Yale University). All oligonucleotides were synthesized through automated DNA synthesis using standard phosphoramidite chemistry and purified by 10% denaturing PAGE. DNA concentration was measured at 260 nm using any UV-vis spectrophotometer, such as NanoVue (GE Healthcare, UK). The sequences of all oligonucleotides are listed below.

Table 4. Sequences used for *F. nucleatum* Selection.

Name	Sequence
Library (K1)	5′- CTAATCTCTACACAGCAGCG -N_{50}-CGTAGTTGAGTCTGAGTGCT
Forward Primer (FP)	5′- CTAATCTCTACACAGCAGCG
Reverse Primer 1 (RP1)	5′- AGCACTCAGACTCAACTACG
Reverse Primer 2 (RP2)	5′- AAAAAAAAAAAAAAAAAAAA-S_9-AGCACTCAGACTCAACTACG (S9 is a triethylene glycol linker)
Ligation Template (T1)	5′- TAGAGATTAGCTTGGTAGTG
Fluorogenic Susbtrate (FQ30)	5′- CTATGAACTGACF-rA-QGACCTCACTACCAAG (F=fluorescein-labeled dT, Q=DABCYL-labeled dT, rA= adenine ribonucleotide)
Deep Sequencing Forward Primer (K1-SFP)	5′- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCTAATCTCTACACAGCAGCG
Deep sequencing Reverse Primer (K1-SRP)	5′- CAAGCAGAAGACGGCATACGAGATAATCCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGCACTCAGACTCAACTACG

2.2.2 Enzymes and chemical reagents

Radioactive nucleotides were purchased from Perkin Elmer (Massachusetts, USA). Enzymes were purchased from MBI Fermentas (Burlington, ON, Canada) with the exception of DNA polymerase, which is purchased from Biotools (Biotools, Spain). Proteinase K was purchased from Thermo Scientific (Massachusetts, USA). Non-radioactive nucleotides are purchased from Fermentas. All other chemicals are purchased from Sigma-Aldrich (Missouri, USA) unless specified otherwise. All multi-component solutions are made in-house using autoclaved, deionized water, filtered through the Milli-Q® Intergral Water Purification System (EMD Millipore, Massachusetts, USA).

2.2.3 Cell lines and cell culture reagents

Fusobacterium nucleatum subsp. nucleatum (ATCC 25586), *Fusobacterium nucleatum subsp. polymorphum* (ATCC 10953), *Fusobacterium necrophorum subsp. necrophorum* (ATCC 25286), *Bacteriodes fragilis* (ATCC 25285) were purchased from American Type Culture Collection (Virginia, USA). *E. coli* K12, *Bacillus subtilis*, *Clostridium difficile* NAP1 strain were routinely cultured and maintained in our laboratory. *Streptococcus salivarius, Collinsella aerofaciens, Bacteroides intestinalis, Bacteroides vulgatus, Dorea longicatena, Coprococcus comes* were generously donated from Dr. Michael Surette lab from the Farncombe Family Digestive Health Research Institute at McMaster University (Hamilton, ON, Canada).

Chopped meat glucose medium was purchased from Sigma-Aldrich while Luria Bertani (LB) Medium was purchased from BioShop (Burlington, ON, Canada). Tryptic soy borth was purchased from BD Biosciences (Mississauga, ON, Canada).

2.2.4 Preparation of crude extracellular mixture (CEM)

Fusobacterium strains: 5 mL of pre-reduced liquid chopped meat broth was inoculated with *F. nucleatum subsp. nucleatum*, *F. nucleatum subsp. polyphorum* or *F. necrophorum subsp. necrophorum* from a frozen stock and grown 37ºC for ~48 hr until OD_{600} reached ~0.8 under anaerobic condition (gas mixture of 80% N_2, 10% CO_2, 10% H_2) in a Whitley Anaerobic Workstation (Don Whitley Scientific, UK). The bacterial culture was transferred into new microcentrifuge tubes and centrifuged at 5,000 g at 4ºC for 10 min. The supernatant, hereinafter referred to as CEM, was recovered and passed through a 0.22 μm syringe filter. All samples were aliquoted into microcentrifuge tubes, flash frozen and stored at -80ºC until further use.

Other bacteria: *E. coli* K12, and *B. subtilis* were cultured using liquid LB broth under aerobic condition until OD_{600} reached ~1. *C. difficile* NAP1, *B. fragilis, S. salivarius, C. aerofaciens, D. longicatena* and *C. comes* were cultured in liquid chopped meat broth under anaerobic condition as described above until OD_{600} reached ~1. *B. intestinalis* and *B. vulgatus* were cultured in liquid tryptic soy broth under anaerobic condition as described above until OD_{600} reached ~0.7. All bacterial CEM was extracted as described above.

2.2.5 In vitro selection of *F. nucleatum* responsive DNAzyme

Ligation: 1 nmole of the initial DNA library (K1) was phosphorylated with 10 μCi of [γ-^{32}P]ATP and 10 units of T4 polynucleictide kinase (PNK) in 1× PNK buffer A (50 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM DTT, and 0.1 mM spermidine, pH 7.6) in a reaction volume of 100 μL and incubated at 37 °C for 15 mins to allow radioactive phosphorylation of K1. Then, nonradioactive ATP was added to a final concentration of 1 mM and incubated for an additional 20 mins at 37 °C. The reaction was quenched by heating at 90°C for 5 min. Equimolar of the substrate (FQ30) and ligation template (T1) were then added and heated 90°C for 1 min and allowed to cool to room temperature for 10 mins. Then, the ligation reaction was initiated by adding 10 units of T4 DNA ligase in 1× T4 DNA ligase reaction buffer (40 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, and 0.5 mM ATP, pH 7.8) in a total volume of 200 μL and incubated at room temperature for 2 hr. DNA was then recovered by ethanol precipitation and separated on a 10% dPAGE. The radioactive PAGE gel was exposed to phosphor imaging plate for 5 mins and imaged on a Typhoon 9200 Imager (Amersham Biosciences, UK) and the gel piece containing the ligated library was eluted and ethanol precipitated.

Counter selection: Ligated DNA library was dissolved in 150 μL of ddH_2O and incubated with 150 μL of 2× selection buffer (100 mM HEPES, 300 mM NaCl, 30 mM $MgCl_2$, and 0.02% Tween 20, pH 7.5) at room temperature for 2 hr. The reaction was quenched by ethanol precipitation and the sample was loaded on a 10% dPAGE to allow the separation of the cleavage product and the uncleaved library. A selection marker was made by artificially inducing RNA cleavage using NaOH. Roughly 10% of the ligated

library was mixed with 5 μL of 0.25 M NaOH and heated at 90°C for 10 mins to induce cleavage. The selection marker was loaded next to the counter selection sample to indicate the position of cleavage and only the uncleaved library was retained and purified from the gel. The counter selection step was conducted to eliminate any active species that responded to the metal ions in the selection buffer independent from *F. nucleatum*.

Selection: The uncleaved library from counter selection was dissolved in 30 μL of ddH_2O and incubated with 150 μL of 2× selection buffer and 120 μL of *F. nucleatum* CEM for 3 hr at room temperature. The reaction was quenched by ethanol precipitation and separated on a 10% dPAGE as described in counter selection. This time, only the cleaved product was isolated and purified from the gel and the resulting DNA pellet was dissolved in 10 μL of ddH_2O and used for PCR in the next two steps.

PCR1: A PCR mixture containing 0.5 μM FP, 0.5 μM RP1, 200 μM dNTP mix (dGTP, dATP, dCTP and dTTP) and 2.5 units of Biotools DNA polymerase (Biotools, Spain) was mixed with 5 μL of the cleaved library in 1× PCR reaction buffer (75 mM Tris-HCl, 2 mM $MgCl_2$, 50 mM KCl, and 20 mM $(NH_4)_2SO_4$, pH 9.0) in a total reaction volume of 50 μL. PCR reaction is carried out under the following condition: denaturation at 94°C for 4 min; followed by 8-12 cycles (dependent on the amount of DNA template recovered from previous selection step) of denaturation at 94°C for 1 min, annealing at 55°C for 40 s and extension at 72°C for 40 s; and a final extension cycle at 72°C for 1 min. 5 μL of the PCR1 products were loaded on a 2% agarose gel stained with 0.005% (v/v) SYBR safe DNA Gel Stain. The agarose gel was imaged using Typhoon 9200 to confirm successful amplification of the cleaved library.

PCR2: 1 μL of PCR1 product was diluted into 50 μL with ddH_2O. A 50 μL PCR reaction mixture was prepared using primers FP and RP2, and 1 μL of the diluted PCR1 product. All other PCR reagents are the same as described in PCR1 and 12 cycles of amplification was performed using the same PCR condition. In order to obtain ~100 pmol of PCR2 product for the next round of *in vitro* selection, 12 individual 50 μL PCR reactions were carried out concurrently. PCR products were analyzed on a 2% agarose gel to ensure proper amplification. PCR products from all 12 reactions were combined, ethanol precipitated and separated on a 10% denaturing PAGE to isolate the sense strand. Because RP2 contains a 20 nt overhang at the 5′ end separated by a triethylene glycol linker, the overhang cannot be amplified by DNA polymerase. This makes the antisense strand 20 nt longer than the sense strand. Therefore, the bottom band on the PAGE gel was excised and eluted to purify the sense PCR product. By the end of PCR2, ~ 100 pmol of amplified DNA was obtained and used for the next round of selection.

Selection cycle: In total, 9 rounds of *in vitro* selection were performed by repeating the protocols described above. Counter selection against selection buffer was performed for round 1, 2, 3, 4, 6, 7 and 9 at 2 hr. Positive selection for *F. nucleatum* was conducted for 3 hr at round 1, 2 hr at round 2 and 1 hr for all subsequent rounds. After the initial ligation, ~ 100 pmol of amplified DNA library was ligated to equimolar amount of FQ30 for all subsequent rounds while other procedure remained the same.

2.2.6 Sequencing analysis

Sequencing: The cleavage product from the final round of selection was amplified by PCR to obtain sufficient DNA for sequencing. PCR1 was conducted using FP1 and RP1 following the same protocol as described above. 1 μL of the PCR1 product was diluted into 100 μL with ddH_2O and 2 μL was used as the template for PCR2 using deep sequencing primers K1-SFP and K1-SRP following the same protocol above for PCR1. 8 individual PCR reactions were performed and the PCR products were purified by 2% agarose gel electrophoresis. DNA extraction from agarose gel was done using GenElute Gel Extraction Kit (Sigma Aldrich). Purified PCR products were sequenced using paired-end next generation sequencing (NGS) using an Illumina Miseq system at the Farncombe Metagenomics Facility, McMaster University.

Sequence analysis: Raw sequencing reads were first trimmed of their primers using Geneious. The resulting 50 nt reads were filtered for quality using PrinSeq v0.20.4 to make sure only high quality reads were used for further analysis.[76] All sequences with any bases of Phred scores < 20 (base-call probability < 99%) were eliminated. Using a clustering algorithm CD-HIT-EST, sequences were grouped into clusters (also known as classes).[77] The following input parameters were used: identity threshold (-c), 0.9; word length (-n), 7; (-d), 0; (-g), 1. Grouped classes were then ranked by size, defined by the number of sequences in that class, to identify the dominating sequences in the pool.

2.2.7 DNAzyme activity assay

DNAzyme RFD-FN1 constructs (DT1 to DT4) were synthesized through standard phosphoramidite chemistry and purified by 10% denaturing PAGE. Cleavage reactions were set up by mixing 23 μL of CEM-FN with 1 μL of 0.25 μM substrate (S) and 1 μL of 25 μM RFD-FN1 constructs. For negative control, 23 μL of ddH_2O was added instead of CEM-FN. Reactions were initiated by adding 25 μL of 2× SB to all samples and incubated at room temperature for 4 hr. Reactions were quenched by adding 1 μL of 0.5 M EDTA (pH 8.0), 5 μL of sodium acetate (3 M, pH5.2) and 135 μL of 100% ethanol for DNA precipitation. After ethanol precipitation, cleavage bands were separated by 10% dPAGE, visualized by Typhoon 9200 and quantified by Image Quant 5.2.

2.2.8 Optimization of reaction conditions

Varying divalent metal ions: Stocks of 2× SB containing 100 mM HEPES, pH 7.5, 300 mM NaCl, 0.02% Tween 20 and 30 mM MCl_2 (where M = Cd, Co, Mg, Mn, Ni, Cu, Zn and Ca) were prepared. Cleavage activity of RFD-FN1 DT4 was assayed according to the description in Section 2.2.7 using CEM-FN and 2× SB containing various divalent metal ions. Cleavage activity was analyzed by 10% dPAGE, visualized by Typhoon 9200 and quantified by Image Quant 5.2.

Varying Mg concentration: Stocks of 2× SB containing 100 mM HEPES, pH 7.5, 300 mM NaCl, 0.02% Tween 20 and various concentrations of $MgCl_2$ ranging from a final $MgCl_2$ concentration of 0, 1, 5, 7.5, 10, 15, 20, 30 and 50 mM. Cleavage activity of RFD-FN1 DT4 was assayed as described above.

Varying pH: A series of 2× SB stock containing 300 mM NaCl, 30 mM $MgCl_2$, 0.02% Tween 20 and 100 mM of a chosen buffering agents with varying pH that ranges from 5.0 to 9.0 at an increasing interval of 0.5 units. MES was used for pH 5.0, 5.5 and 6.0; HEPES was used for pH 6.5, 7.0, 7.5 and 8.0; Tris was used for pH 8.5 and 9.0. The cleavage reaction of RFD-FN1 DT4 was conducted using CEM-FN as described above.

Varying Reaction Temperature: Cleavage analysis of RFD-FN1 DT4 at different incubation temperature was conducted using CEM-FN in 2× SB (100 mM HEPES, 300 mM NaCl, 30 mM $MgCl_2$, and 0.02% Tween 20, pH 7.5) in the same experimental setup as described in Section 2.2.7. Each reaction was incubated at 4, 15, 20, 23, 37 and 65°C for 4 hr. Cleavage of RFD-FN1 DT4 in selection buffer alone without CEM-FN was also performed at the corresponding temperatures to control for the background cleavage of RNA at higher temperatures. Cleavage activity at each reaction temperature was calculated by subtracting the cleavage in selection buffer alone from the cleavage in CEM-FN. All reactions in this assay were done in triplicates.

2.2.9 Comparison of the cleavage activity of RFD-FN1 in CEM-FN and CIM-FN

CIM-FN (the crude intracellular mixture of *F. nucleatum*) was produced as follows: *F. nucleatum* cells were grown exactly the same as described above. 1 mL of bacterial culture was centrifuged at 5,000 g at 4ºC for 10 min. Instead of taking the supernatant, the cell pellets were retained and re-suspended in 150 µL of 1× SB, heated at 90ºC for 15 min and centrifuged at 11,000 g at 4ºC for 5 min. The resulting supernatant, hereinafter referred to as CIM-FN, was aliquoted and stored at -80ºC until further use.

For comparison, either CEM-FN or CIM-FN was used to induce the cleavage activity of RFD-FN DT4 construct. The cleavage reactions were set up exactly the same way as described above.

2.2.10 Heat resistance analysis

To evaluate the thermal stability of the potential targets, CIM-FN was subjected to further heating at 90ºC for 5, 10, 20, 30 or 60 min. The cleavage activity of DT4 was induced by adding CIM-FN heated for different times in the same assay set up as described above.

2.2.11 Potential target analysis

To treat CIM-FN with protease digestion, 1 µL of 20 mg/mL proteinase K was mixed with 23 µL of CIM-FN and 25 µL of 2× SB. The resulting mixture was incubated at 50ºC for 1 hr. Afterwards, 1 µL of 0.25 µM substrate and 1 µL of 25 µM DT4 were added to the mixture and the cleavage reaction was carried out at room temperature for 4 hr. Cleavage analysis was performed as described above.

To estimate the molecular weight of the potential protein target, CIM-FN was filtered through different molecular sizing centrifugal columns. 200 µL aliquots of CIM-FN were individually filtered through a 3 kD, 10 kD, 30 kD, 50 kD or 100 kD columns (Nanosep Omega, Pall Incorporation). 23 µL of the filtrates from each column were used to induce the cleavage activity of DT4 and 10% dPAGE was used to analysis.

2.2.12 Specificity analysis

The bacterial strains listed in Section 2.2.3 were cultured according to the procedures described in Section 2.2.4. The CIM of each bacterium was isolated according to Section 2.2.8 and 23 μL of each CIM was used to induce the cleavage activity of DT4, followed by 10% dPAGE analysis.

2.2.13 Detection sensitivity analysis

5 mL of pre-reduced liquid chopped meat broth was inoculated with *F. nucleatum* from a frozen stock and grown for ~48 hr until OD_{600} reached ~0.7 under anaerobic condition. The bacterial culture was serially diluted in 10-fold intervals 10 times. Triplicates of 100 μL of each dilution were plated onto Brucella Blood Agar plates to establish CFU/mL cell counts. At the same time, 1 mL culture from each dilution was harvested and the corresponding CIM was produced to induce the cleavage activity of DT4 and analyzed by 10% dPAGE. This experiment was used to assay the detection sensitivity of DT4 without additional culturing.

To establish the detection sensitivity with various culturing times, 11 culture tubes containing 5 mL of pre-reduced chopped meat broth were set up and each inoculated with 100 μL of each serial dilution. Each tube was allowed to grow under anaerobic condition at 37ºC and a 1-mL solution was harvested from each tube at 12, 24 and 36 hr. The corresponding CIM was processed to induce the cleavage activity of DT4, followed by 10% dPAGE analysis.

2.2.14 Fecal spiking assay

50 mg of fecal sample from healthy patient #3 was added to 1 mL of culture *F. nucleatum* cells (roughly 10^8 cells). The mixture was vortexed vigorously and passed through a 40 μm cell strainer (Corning Inc, New York, USA) to remove any remaining solid particles. Filtrates were spun down at 5,000 g for 10 min to collect bacterial cells. The supernatant was removed while the cell pellet was resuspended in 100 μL of 1× SB and heated at 90ºC for 15 min. The cell debris was spun down at 14,000 g and the supernatant was retained. 23 μL of the resulting supernatant was used to induce DT4 DNAzyme assay as described above.

Chapter 3. Results and Discussions

3.1 In vitro selection of CRC-responsive RFD

3.1.1 Selection-CRC1 results

In vitro selection of an RNA-cleaving DNAzyme directly using human fecal samples is innovative, but also experimentally challenging due to the complexity of human biological samples. An initial attempt at isolating a CRC-specific RFD conducted by Dr. Weijia Zhu showed that it was possible to enrich sequences that were cleaved in the presence of CRC fecal samples (data not shown). However, the enriched sequences were not cleaved at the putative RNA cleaving site. Sequence cleavage at any position other than the RNA would yield a poor fluorescent signal enhancement due to the placement of the fluorophore and quencher pair flanking the RNA. Therefore, the lack of cleavage specificity rendered these sequences unsuitable for further analysis. Nonetheless, the preliminary data showed strong indications that with the right positive and counter selection strategy it is possible to enrich DNAzymes that are responsive to targets within CRC fecal samples.

In a second selection experiment (herein referred to as selection-CRC1), a stringent selection strategy was devised. All fecal samples from 15 CRC patients and 4 healthy controls were pooled into four groups (CRC Mix A, CRC Mix B, CRC Mix C and Healthy Mix) according to Table 3. This accounts for the chemical and biological heterogeneity expected from fecal samples that come from different individuals.[4, 14] CRC Mix A, B & C were used as targets for each round of selection in alternation. This was designed to guide potential DNAzymes to target a molecule that is shared among all

CRC patients. To ensure target specificity, counter-selection was introduced in-between positive selections using healthy control samples. Overall, 7 rounds of positive selection and 1 round of counter-selection were completed. The calculated percent cleavage at each round of selection is illustrated in Figure 6. During the 1st round, a 12.8% cleavage was observed as a result of the 2 nmol of DNA used as the initial library. Round 2 to 7 showed no significant increase in cleavage as the cleavage rate fluctuated around 2%-6%.

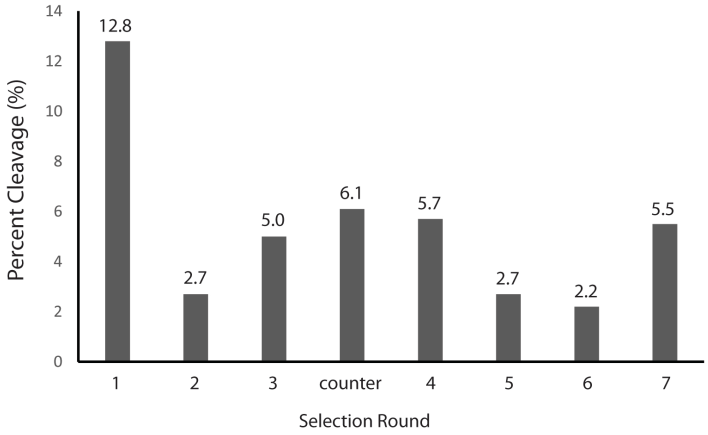

Figure 6. *Selection-CRC1 results***.** In this selection, CRC fecal samples were pooled into groups of mix A, mix B and mix C as the positive selection target. All four healthy control samples were grouped together for negative selection. For every round of selection, the DNA pool was ligated to the FQ30 substrate and incubated with the target in 1× selection buffer at room temperature for various times. Each round of selection was purified by dPAGE and the percent cleavage at each round was quantified by ImageQuant 5.2.

Deep sequencing results from round 4 revealed a total of 165,073 sequences that were grouped into 161,759 classes based on the sequence similarity. These classes were ranked by their abundance and the top 5 classes are summarized in Table 5. Based on the ratio of total number of classes over total number of sequences, limited sequence enrichment was observed at round 4. This was also evident by the abundance of the top 5

classes, which all had less than 30 sequences, with the highest frequency of only 0.0127%. At round 7, the level of enrichment of catalytic sequences improved slightly over round 4 as indicated by the increased frequency of the top classes.

Table 5. Sequencing analysis of Selection-CRC1.

	Round 4		Round 7	
Total Sequences	165,073		23,203	
Total Clusters	161,759		14,622	
Class	Copy Number	Frequency (%)	Copy Number	Frequency (%)
1	21	0.0127	105	0.4525
2	16	0.0097	100	0.4309
3	14	0.0085	75	0.3232
4	5	0.003	59	0.2542
5	4	0.0024	56	0.2413

3.1.2 Characterization of CRC1 top sequences

Next, the top 3 sequences from both round 4 and round 7 were tested for their cleavage activity in the presence of CRC patient fecal samples as well as healthy controls. As shown in Figure 7-A, class 1 sequence from round 4 (named C4-1) showed stronger cleavage activity towards CRC mix A and mix B compared to the healthy mix, with cleavage activities of 28.5%, 33.5% and 9.7% respectively. However, a significant degree of nuclease degradation was observed based on the smearing and the nonspecific banding patterns on the dPAGE gel. To solve this problem, heating of fecal extracts at 90°C for 10 min was tested. As shown in Figure 7-B, heating the targets eliminated nuclease degradation, but also completely abolished cleavage activity in both CRC samples and healthy controls. This suggests that the observed cleavage in Figure 7-A was either

caused by nuclease-mediated cleavage or that the target was heat sensitive. In addition, reduction of the quantity of target used in the reaction from 10μL to 3μL significantly reduced cleavage activity to 5.8% and 5.2% in CRC and healthy samples respectively (Figure 7-C). Likewise, decreasing the reaction time from 1 hr to 20 min also reduced cleavage activity drastically as shown in Figure 7-D. Altogether, the observed cleavage of C4-1 in the presence of CRC fecal mix and healthy mix was likely caused by nuclease degradation. Consequently, this sequence is not suitable for further analysis. The top 2 and top 3 classes from round 4 and all three classes from round 7 were tested in the same condition as C4-1. All the sequences showed nuclease-mediated cleavage in both CRC samples and healthy controls. Therefore, results from the selection-CRC1 experiment demonstrates that there is a substantial level of nuclease degradation in complex human fecal samples, and that it interferes with the selection of a true RNA-cleaving DNA enzyme.

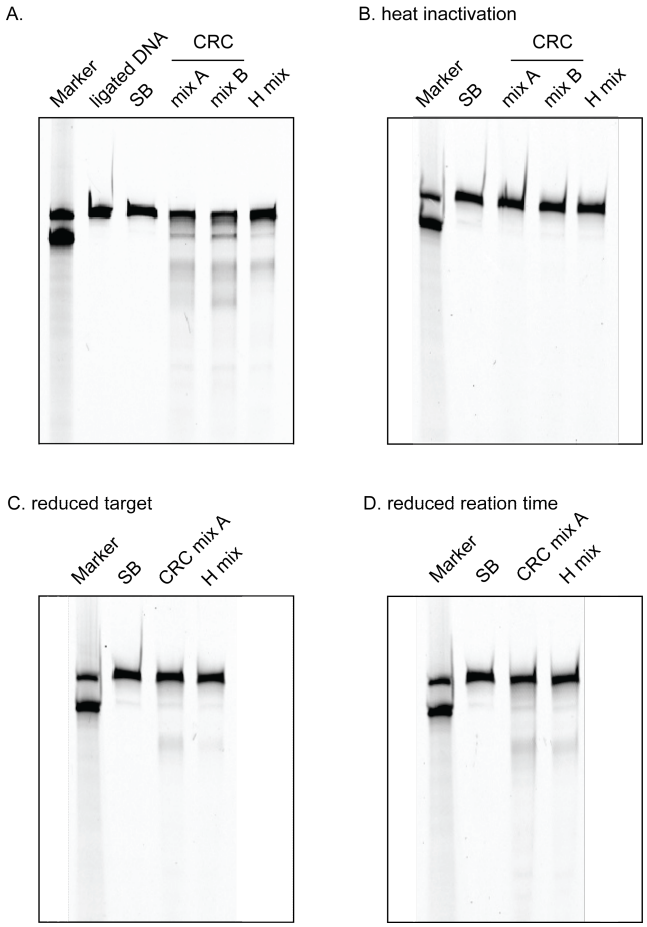

Figure 7. *Characterization of round 4 class 1 sequence.* (A). The catalytic activity of Class 1 sequence from round 4, C4-1, was tested. The sequence was ligated to FQ30 and incubated with CRC mix A, mix B and healthy mix for 1 hr at room temperature in selection buffer containing 300 mM NaCl, 100 mM KCl, 10 mM $MgCl_2$, 5 mM $MnCl_2$, 60 mM HEPES, pH 7.5, 0.01% Tween 20. (B). Fecal samples were heat inactivated at 90°C for 10 min and tested with C4-1. (C). Target quantity was reduced from 10 µL to 3 µL. (D). The reaction time was reduced from 1 hr to 20 min. Marker represents NaOH treated cleavage. SB represents the incubation with selection buffer only.

3.1.3 Nuclease degradation analysis

Oligonucleotides are prone to nucleolytic degradation in bodily fluids such as serum and feces. This has presented a great challenge for the selection of DNAzymes against complex human fecal samples because the DNA library is constantly being non-specifically degraded by nucleases during each round of selection. In the past, chemically modified nucleotides have been applied extensively in antisense technology to enhance biostability. Yang *et al.* showed that 2′-O-methyl modification of a ribozyme at 31 out of the 37 positions led to a 1000-fold increase in nuclease resistance over that of the all-RNA ribozyme while retaining its catalytic activity.[78] Addition of a 3′-3′ linked, inverted thymidine at the 3′-end of oligonucleotides has also been shown to inhibit the degradation of DNA by 3′-exonucleases.[79] Similarly, phosphorothioate bonds have been used widely to increase the stability of DNAzymes.[80] Among the available chemical modifications, phosphorothioate bond was further investigated due to its inexpensive synthesis and compatibility with the in vitro selection process.

In order to evaluate the effect of having phosphorothioate bonds as a way to inhibit nuclease degradation, a 5′-phosphorothiolated polyT sequence was ligated at the 5′-end of the FQ30 substrate, which is then ligated to the class 1 sequence from round 4. Both the modified full-length sequence and the original sequence were incubated with one CRC fecal sample at room temperature for 1 hr. As shown in Figure 8-B, 5′-phosphorothioate modification produced a different degradation pattern on the gel, but did not reduce the overall level of nuclease digestion compared to the unmodified sequence. In further analysis, the same 5′-phosphorothiolated polyT sequence with FQ30

was ligated to the 5′-end of the original DNA library. A time-trial experiment with the modified library showed that it was completely degraded within 2 hr while the unmodified library lasted approximately 4 hr (Figure 8-C). Therefore, the 5′-phosphorothiolated-polyT sequence used here did not provide any protection against nuclease degradation. Because fecal samples contain both 5′- and 3′-exonucleases and endonucleases, it is possible that combinations of chemical modifications are required to fully eliminate the nuclease problem.

Taken together, no further incorporation of any chemical modification were made to future selections because the addition of modified nucleotides adds an extra level of difficulty on an already difficult selection of a complex biological sample. For simplicity, the samples were heated to eliminate any nuclease activity. A previous selection from our lab demonstrated that the RFD-CD1 isolated from the CEM of *Clostridium difficile* was specifically targeting a heat resistant protein, TcdC.[66] This suggests that there are possible heat resistant proteins present in human fecal samples that can be isolated through in vitro selection.

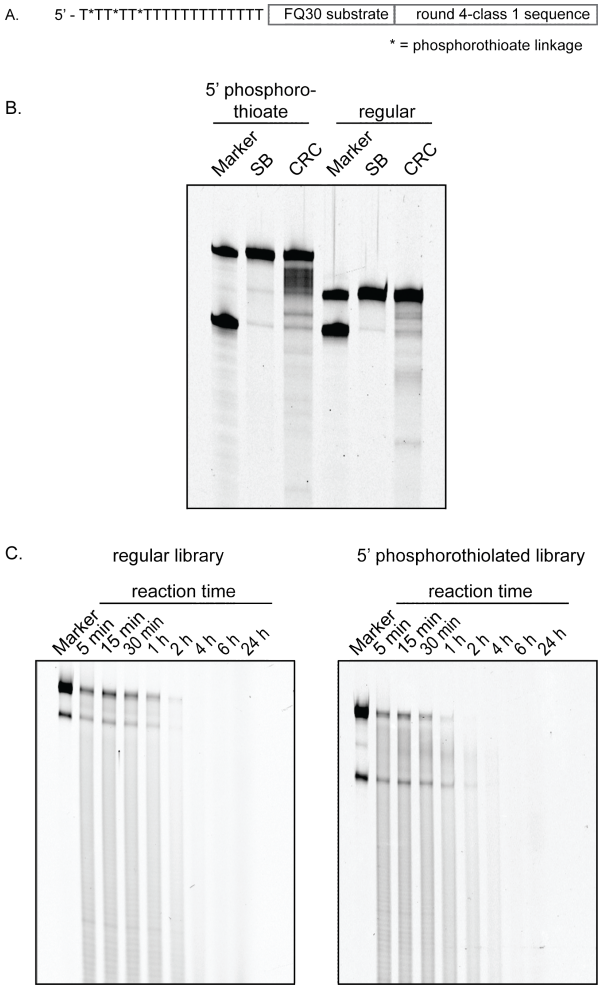

Figure 8. *Nuclease degradation analysis.* (A). An 18 nucleotide polyT sequence containing 3 phosphorothioate linkages at the 5′ end was ligated to FQ30 at the 5′ end and further ligated to C4-1. (B). Both the ligated phosphorothiolated sequence and the unmodified sequence were tested with one CRC patient fecal sample (P7) for 1 hr at room temperature. (C). The 5′ phosphorothioate modification was ligated to the DNA library. Both the regular library and the modified library were tested with one CRC fecal sample (P7) for up to 24 hr in order to evaluate the degree of nuclease degradation.

3.1.4 Selection-CRC2 with heat denaturation

In this selection experiment (referred to as selection-CRC2), all fecal samples were heated at 90°C for 10 min and cooled to room temperature before being used in the selection reaction. The selection strategy was simplified by combining all CRC patient fecal samples as one group and all the healthy samples together as the control. In order to prevent the loss of catalytic sequences too early, counter selection was not performed until a significant enrichment of catalytic sequences was observed. When the activity of an RNA-cleavage is evident, we can begin to integrate counter selection to control specificity. Overall, 7 rounds of positive selections were completed and the calculated percent cleavage at each round of selection is illustrated in Figure 9.

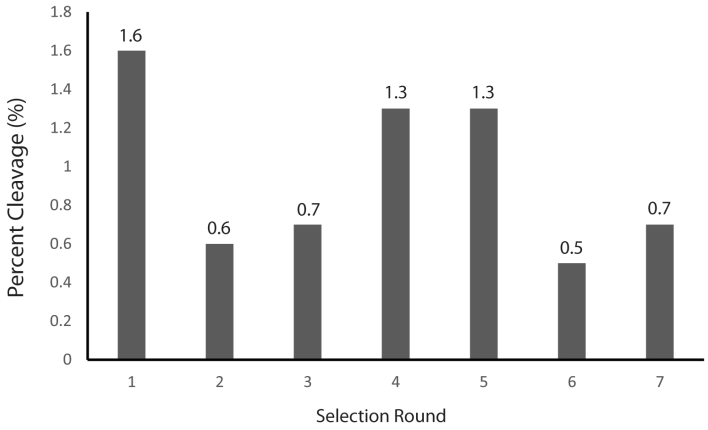

Figure 9. *Selection-CRC2 results.* For the revised selection, all CRC fecal samples were pooled together as the positive selection target. In each round, the target was heated at 90°C for 10 min, cooled to room temperature and then incubated with the ligated DNA pool in 1× selection buffer at room temperature. The selection buffer contained 300 mM NaCl, 100 mM KCl, 10 mM $MgCl_2$, 5 mM $MnCl_2$, 60 mM HEPES, pH 7.5, 0.01% Tween 20. Each round of selection was purified by PAGE and the percent cleavage at each round was quantified by ImageQuant 5.2.

Deep sequencing results from round 4 showed that there were 234,960 sequences in the DNA pool, which are grouped into 230,638 classes (Table 6). Once again, there was little enrichment of catalytic sequences at round 4. However, catalytic sequences significantly accumulated after round 7, which had 5,031,868 sequences that were grouped into roughly 1,600,000 classes. The class 1 sequence from round 7 dominated the DNA pool at 44.16%. The top three classes from round 7 were synthesized and further evaluated for their catalytic activities.

Table 6. Selection-CRC2 sequencing analysis.

	Round 4		Round 7	
Total Sequences	234,960		5,031,868	
Total Clusters	230,638		~1,600,000	
Class	Copy Number	Frequency (%)	Copy Number	Frequency (%)
1	42	0.0178	2,221,909	44.1567
2	40	0.0170	31,556	0.6271
3	17	0.0072	11,030	0.2192
4	17	0.0072	5,373	0.1068
5	16	0.0068	1,620	0.0322

3.1.5 Characterization of Selection-CRC2 Top Sequences

All three classes of sequences were tested with selection buffer, CRC mix or healthy mix at room temperature for 1 hr. The class 1 sequence, which dominated the DNA pool at 44.16%, showed significant cleavage activity at 64.1%, 70.8% and 68.5% in buffer, CRC mix and healthy mix respectively (Figure 10). Since the rate of cleavage for all three tests were approximately the same, it suggests that the cleavage is mediated by selection buffer rather than the fecal samples that were used as the target. Similarly, both

class 2 and class 3 sequence showed cleavage in buffer, indicating that they were also buffer-mediated RNA-cleaving DNAzymes. Since most previously isolated RNA-cleaving DNAzymes in literature depend on divalent metal ion for catalysis, it is likely that these RFDs respond either to the Mg^{2+} or Mn^{2+} ions that are present in the selection buffer. Although it is necessary to keep metal ions in the selection buffer for DNAzyme catalysis, certain potent metal ions such as Mn^{2+} tend to induce buffer-mediated cleavage. Therefore, Mn^{2+} ion concentration should be reduced or eliminated for future selections.

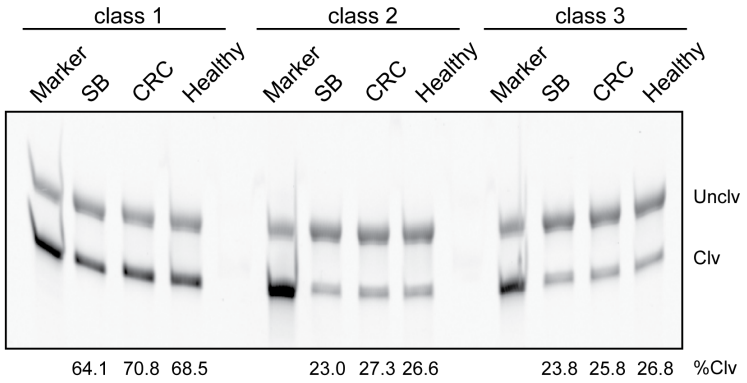

Figure 10. *Characterization of top 3 classes of DNAzymes from selection-CRC2.* (A) The catalytic activity of top three classes from round 7 was tested. All sequences were ligated to FQ30 and incubated with the CRC mix and the healthy mix for 1 hr at room temperature in selection buffer containing 300 mM NaCl, 100 mM KCl, 10 mM $MgCl_2$, 5 mM $MnCl_2$, 60 mM HEPES, pH 7.5, 0.01% Tween 20. Marker represents NaOH treated cleavage. SB represents the incubation with selection buffer only. The percent cleavage rates of the three classes were calculated using ImageQuant 5.2.

3.1.6 Project summary

Colorectal cancer is a serious condition and individuals affected could greatly benefit from a simple, reliable, and non-invasive diagnostic test. The project described here takes a bold step forward to reach such a goal by applying in vitro selection to highly complex biological samples. However, the complexity of crude human fecal matrix has presented many challenges for us. First, there is a significant level of non-specific degradation caused by nucleases that are ubiquitous in human stool samples. This interferes with sequence enrichment during in vitro selection because the method depends on RNA-cleavage as a selectable marker. Unfortunately, the use of chemically modified nucleotides was not effective in reducing the level of nuclease degradation. In our second attempt in selection-CRC2, heat inactivation of the selection target successfully abolished all nuclease activity. However, due to the lack of counter selection against the selection buffer, DNAzymes that were responsive to the metal ions in the selection buffer were isolated. This result has prompted us to include stringent counter selection steps in all future selection experiments.

3.2 In vitro selection of a *Fusobacterium nucleatum*-specific DNAzyme

3.2.1 In vitro selection results

F. nucleatum-specific RFDs were isolated using a 50-nucleotide randomized DNA library via the selection strategy illustrated in Figure 5. Prior to selection, *Fusobacterium nucleatum subsp. nucleatum* was cultured anaerobically to an OD_{600} reading of approximately 0.7 and the cells were removed by centrifugation. The resulting CEM, herein referred to as CEM-FN, was used as the selection target. The DNA library was first ligated to the fluorogenic substrate (FQ30) and incubated in selection buffer for 2 hr as the counter selection step to eliminate ligand-independent RFDs. The uncleaved DNA molecules were purified and incubated with CEM-FN in 1× SB for 3 hr. This served as the positive selection step where RFDs that are responsive to the presence of CEM-FN were isolated. The resulting cleaved sequences were purified by 10% denaturing PAGE, amplified by PCR and used for the subsequent round of selection.

Overall, 9 cycles of selective amplification were completed. In the 9^{th} round, a strong cleavage activity was observed in the presence of CEM-FN relative to the negative control, which contains only the selection buffer (Figure 11-B). The DNA pool obtained from round 9 was sequenced using the Illumina MiSeq platform. Sequencing reads were clustered into classes based on sequence similarity and ranked by sequence abundance. The top five classes of RFDs were analyzed for their catalytic activity and the class 3 sequence, herein referred to as RFD-FN1, displayed the highest cleavage activity and was chosen for further investigation (Figure 11-C).

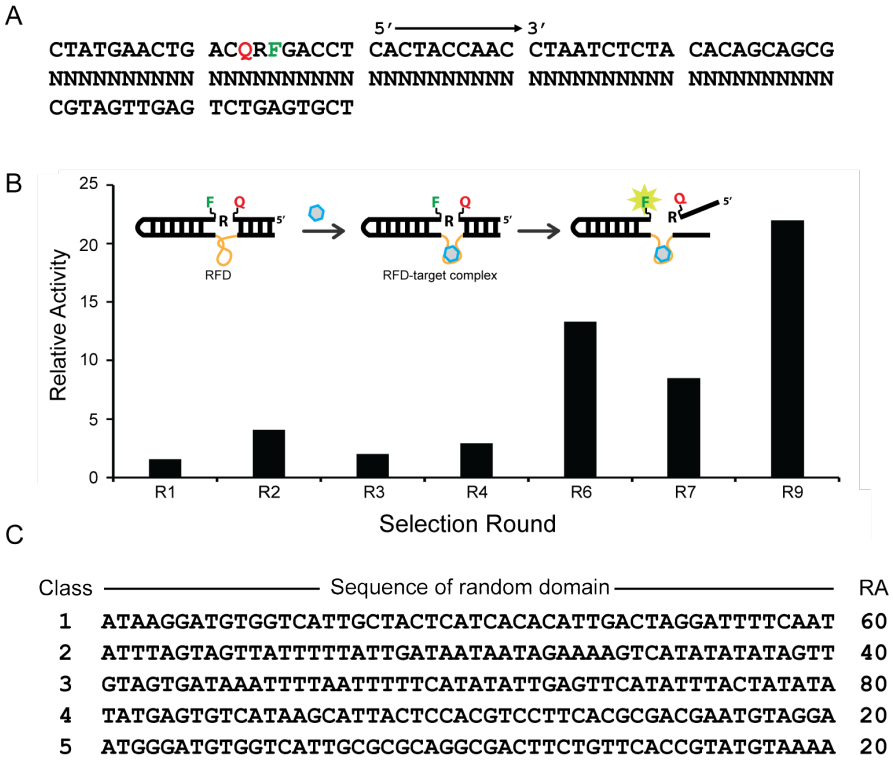

Figure 11. *In vitro selection of Fusobacterium nucleatum-responsive RFDs.* (A) The sequence of the DNA library used in this study. R: adenosine ribonucleotide; F: fluorescein-labeled dT; Q: DABCYL-labeled dT; N: mixture of 25% A, 25% G, 25% C, 25%T. (B) Performance of evolving DNA pools. The relative activity is calculated as (% cleavage in CEM-FN)/(% cleavage in selection buffer only). (C) The random sequence regions of the top 5 classes. The class-3 sequence is named RFD-FN1. RA: relative activity.

3.2.2 Sequence and structure properties of RFD-FN1 constructs

The original *cis*-acting RFD-FN1 DNAzyme was isolated to cleave a covalently attached substrate sequence. However, a *trans*-acting DNAzyme provides additional advantage of ease-of-synthesis because the DNAzyme sequence is no longer linked to the substrate sequence, which removes the ligation step. Therefore, the original *cis*-acting RFD-FN1 was converted into its *trans*-acting form. Several bimolecular constructs of

RFD-FN1 (DT1 to DT4) were generated where the substrate domain (nucleotides 1 to 30) was detached from the catalytic sequence of RFD-FN1 (nucleotides 31 to 120), as shown in Figure 12-A. DT1 represents the full-length catalytic sequence while DT2 to DT4 represent three different truncations. The shortest construct, DT4, where the DNAzyme strand is made of the first 38 random-sequence nucleotides of RND-FN1, displayed an improved CEM-FN dependent activity (16% cleavage in DT4 vs. 5% cleavage in full-length construct DT1).

In-depth analysis of the substrate and DNAzyme sequence of DT4 revealed three possible duplex elements, illustrated as P1, P2 and P3 in Figure 12-B. The proposed secondary structure of DT4 is reminiscent of many reported RNA-cleaving DNAzymes, such as the well-known RNA-cleaving DNAzymes 8-17 and 10-23.[48, 59, 60-62] P1 and P3 serve as substrate-binding arms while P2 acts as a part of the catalytic core. Interestingly, the suspected catalytic core of RFD-FN1 DT4, made of P2 and SS12, is highly rich in A and T (91.7%): out of 24 nucleotides, 14 and 8 are thymine (58.3%) and adenine (33.3%) nucleotides, respectively. To our knowledge, DNAzyme sequences of this nature have never been reported before.

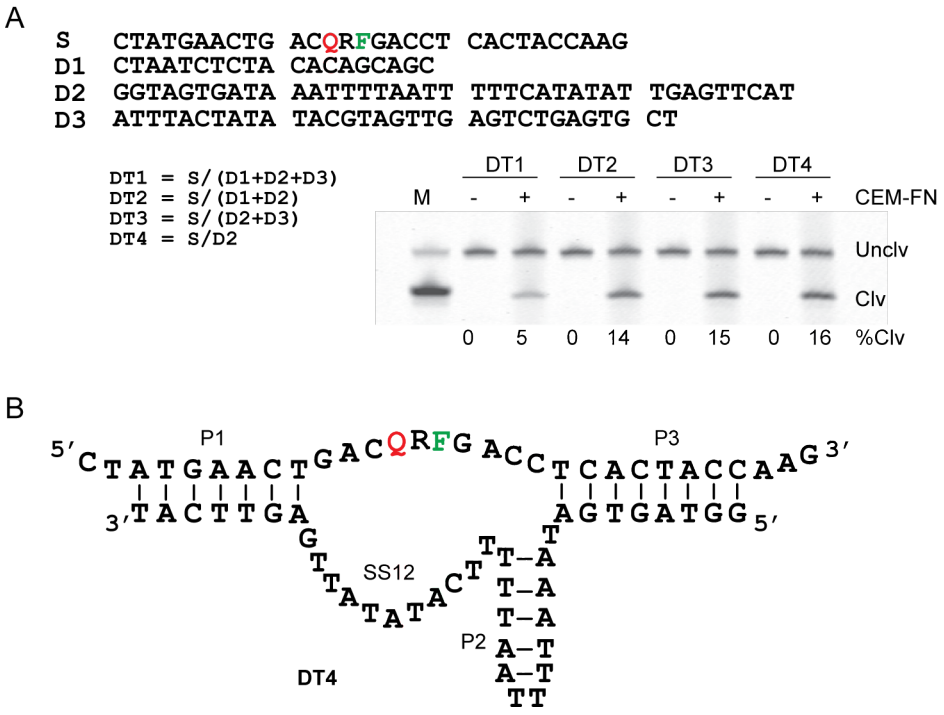

Figure 12. *Sequence and structure properties of RFD-FN1*. (A) Cleavage activities of trans-acting DNAzyme constructs. The DNAzyme strand of RND-FN1 is broken down into three sequence segments: D1, D2 and D3 as shown. DT1 represents the full-length construct (made of all three segments) while DT2 is made of D1 and D2, DT3 is made of D2 and D3, and DT4 is made of D2 only. (B) The proposed secondary structure of DT4 in the presence of the substrate. P: pairing region; SS12: single-stranded motif linking P1 and P2.

3.2.3 Metal ion specificity, pH and temperature dependency of RFD-FN1

Divalent metal ions have been known to play important roles in the activity of RNA-cleaving DNAzymes.[81] In a previous study on the *E. coli* RFD-EC1 DNAzyme, it was demonstrated that RFD-EC1 displayed the highest activity in Ba^{2+} even though it was originally selected using Mg^{2+}.[82] Therefore, the effects of different divalent metal ions on the *F. nucleatum* RFD-FN1 DNAzyme were tested. It was found that RFD-FN1 displayed a strong activity only in Mg^{2+}, which is the metal ion it was selected in

originally. In contrast, Mn^{2+} and Co^{2+} induced weak cleavage (3-4%) while Ca^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+} and Ba^{2+} did not induce any cleavage (Figure 13-A). The effect of different Mg^{2+} concentration on RFD-FN1 activity was also tested and it was found that the highest activity is achieved with a final Mg^{2+} concentration of 10 mM.

Next, we examined the cleavage activity of RFD-FN1 under different pH and reaction temperatures. We found that highest activity was achieved at pH 7.5 and under room temperature (23ºC), both of which were the original reaction condition used during selection.

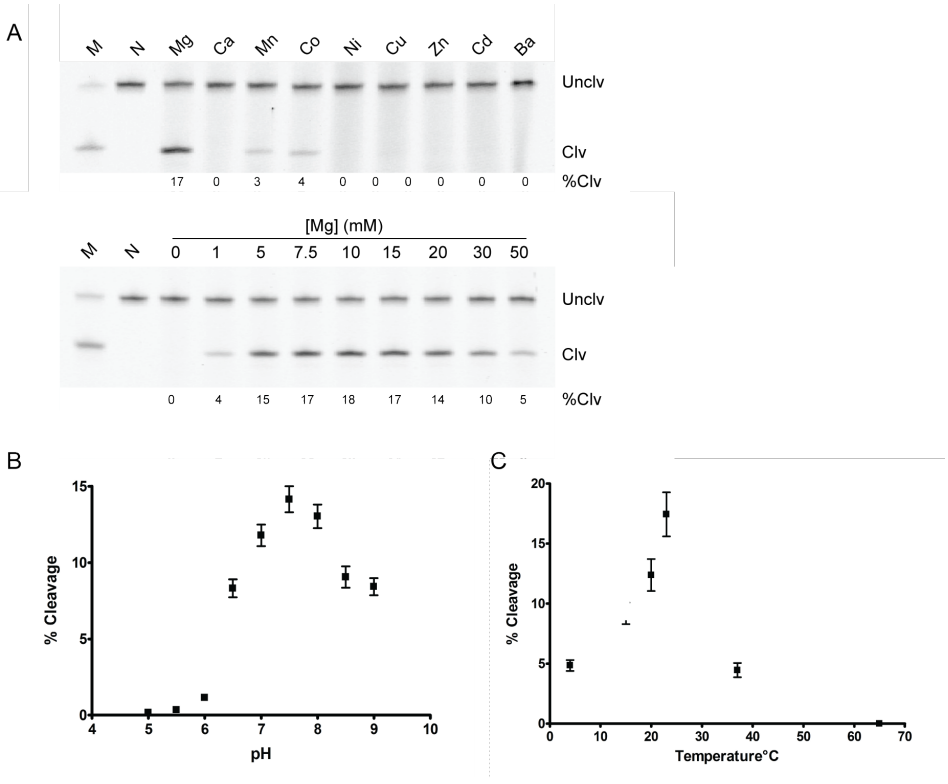

Figure 13. *Characterization RFD-FN1 in varying reaction conditions*. Activity of DT4 in the presence of CEM-FN in various divalent metal ions and Mg concentrations (A), pH (B), and reaction temperature (C). Values in (B) & (C) are the average of three independent experiments.

3.2.4 Characterization of the activator for RFD-FN1

RFD-FN1 was isolated to cleave in the presence of CEM-FN. It is conceivable that the target that activates this DNAzyme also exists inside cells, as previously shown by the RFD-EC1 DNAzyme.[82] This concept was tested by performing the following experiment: *F. nucleatum* cells were first cultured and then pelleted by centrifugation. The supernatant was taken as the CEM-FN. The cell pellet was re-suspended in 1× SB and then heated at 90°C for 15 min. The sample was centrifuged to pellet cell debris and the resulting supernatant was labeled as CIM-FN (CIM stands for crude intracellular mixture). Both CEM-FN and CIM-FN were then used to induce the cleavage activity of DT4 and the results are shown in Figure 14-A. This experiments revealed that the target for RFD-FN1 exists both inside and outside of cells. Remarkably, RFD-FN1 exhibited a high-level activity in CIM-FN, despite the fact that CIM-FN was derived by heating cells at 90°C for 15 min. This observation suggests that the activating target is thermally stable. This was confirmed by the data shown in Figure 14-B where relative activity (RA) was measured after CIM-FN was further treated at 90°C for 5, 10, 20, 30 and 60 min (the activity with no heat treatment was taken as 100). Although CIM-FN further treated at 90°C showed reduced activity, the remaining activity was still quite high. For instance, CIM-FN treated at 90°C for 20 min still induced 55% activity when compared to no heat treatment. This property was further explored below to achieve the detection specificity.

Based on the observation above, it is possible that the DNAzyme activator is a small metabolite as small molecules are in general more stable than macromolecules such as proteins. This hypothesis was refuted when treating CIM-FN with proteinase K for 1 hr

prior to DNAzyme activation completely abolished the cleavage activity of DT4. This suggests that the target is proteinaceous in nature. Fractionation of CIM-FN with molecular weight cut-off filters suggests that the protein target has a molecular weight of approximately 50,000 to 100,000 Daltons. (i.e. cleavage activity was significantly reduced in the 50,000 Daltons filtrate). Currently, the exact identity of the molecular target for RFD-FN1 is under investigation and the results will be reported in the future.

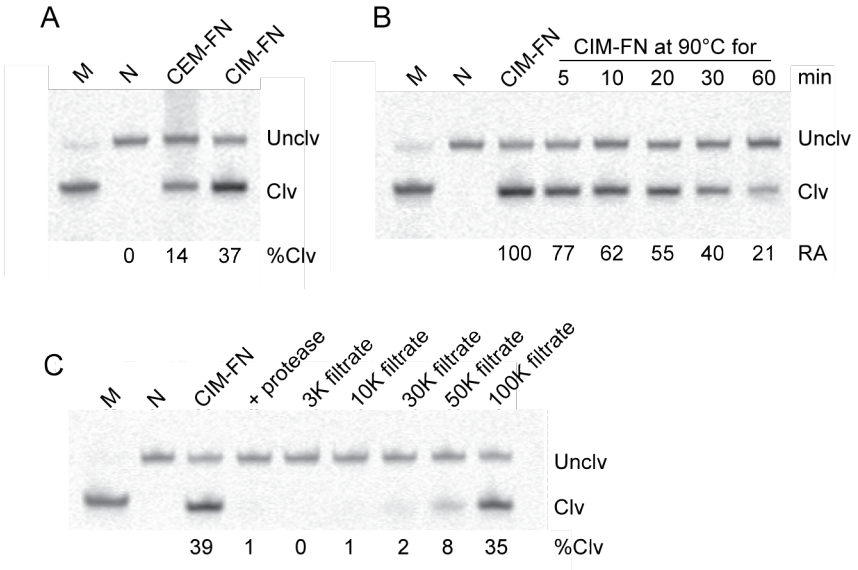

Figure 14. *Characterization of the activator for RFD-FN1.* (A) Activity of DT4 in response to CEM-FN and CIM-FN. (B) Activity of DT4 to CIM-FN treated at 90ºC for different times. RA: relative activity, calculated as C_t/ C_0; C_0 = % cleavage of DT4 in CIM-FN without heat treatment; C_t = % cleavage of DT4 in CIM-FN treated at 90ºC for a given time. (C) Response of DT4 to protease-treated CIM-FN and estimation of the molecular weight of the target. CIM-FN was passed through centrifugal filters with specific molecular weight cut-off individually and tested for cleavage activity.

3.2.5 Recognition specificity of RFD-FN1

Next, the detection specificity of RFD-FN1 was examined. The cleavage activity of DT4 was tested in response to the CIM of several anaerobic bacterial species commonly found in our gut microbiota, as well as two aerobic bacteria. These include: *Bacteriodes fragilis* (BF); *Clostridium difficile* NAP1 (CD); *Bacteroides vulgatus* (BV); *Bacteroides intestinalis* (BI); *Coprococcus comes* (CC); *Dorea longicatena* (DL); *Collinsella aerofaciens* (CA); *Streptococcus salivarius* (SS); *Escherichia coli* K12 (EC, aerobic) and *Bacillus subtilis* (BS, aerobic). DT4 displayed strong cleavage activity only in response to CIM-FN. The CIM of *E. coli* induced a slight cleavage whereas the CIM of *B. subtilis* induced non-specific degradation, likely caused by nucleases based on the smearing pattern on dPAGE. CIMs from the remaining bacteria did not induce any cleavage. In order to eliminate any cross reactivity caused by non-specific nuclease degradation, we took advantage of the heat resistant property of the RFD-FN1 activator. CIMs from all bacteria were additionally heat-treated at 90ºC for 15 min and used to induce cleavage activity of DT4. This completely abolished the faint cleavage activity previously exhibited by *E. coli.* Even though *B. subtilis* still induced non-specific degradation, it did not induce any cleavage at the putative RNA site. Overall, these results suggest that RFD-FN1 is highly specific for *F. nucleatum* under the anaerobic gut environment.

To further understand the level of specificity of RFD-FN1, we examined the cleavage activity of DT4 in response to the CIM of *Fusobacterium necrophorum subsp. necrophorum* (FNn), which is another species under the same genus *Fusobacterium.*

CIM-FNn did not induce any cleavage activity, demonstrating the species-specificity of RFD-FN1. Interestingly, when we tested the cleavage of DT4 towards another subspecies, *Fusobacterium nucleatum subsp. polymorphum* (FNp; note: the subspecies used in the selection is *Fusobacterium nucleatum subsp. nucleatum),* the CIM of FNp induced a slight cleavage. It is possible that RFD-FN1 is activated by the same target in both subspecies of *Fusobacterium nucleatum* and that the level of target is significantly higher in *Fusobacterium nucleatum subsp. nucleatum* than in *Fusobacterium nucleatum subsp. polymorphum.*

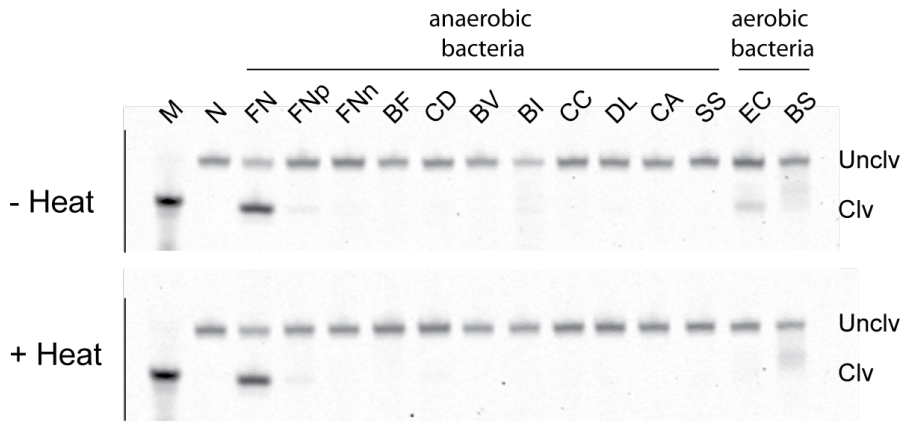

Figure 15. *Recognition specificity of RFD-FN1*. Reactivity of DT4 in the presence of CIM of some common anaerobic bacteria in human gut flora and in the presence of CIM of two common aerobic bacteria. Anaerobic bacteria: FN, *Fusobacterium nucleatum subsp. Nucleatum;* FNp*, Fusobacterium nucleatum subsp. polymorphum;* FNn*, Fusobacterium necrophorum subsp. necrophorum;* BF*, Bacteriodes fragilis;* CD*, Clostridium difficile* NAP1; BV, *Bacteroides vulgatus;* BI*, Bacteroides intestinalis;* CC*, Coprococcus comes;* DL*, Dorea longicatena;* CA*, Collinsella aerofaciens;* SS*, Streptococcus salivarius.* Aerobic bacteria: EC*, Escherichia coli* K12; BS, *Bacillus subtilis.* Heat: CIM additionally heated 90ºC for 15 mins. No heat: CIM without additional heating step.

3.2.6 Detection sensitivity of RFD-FN1

We also investigated the utility of RFD-FN1 as a sensor for *F. nucleatum* detection. Using the gel based method, DT4 was able to achieve a limit of detection of 10^7 CFU/mL (Figure 16-A). However, with the incorporation of a culturing step, the detection sensitivity can be greatly improved. When a 24 hr culturing time is added, RFD-FN1 can detect down to 100 seeding cells and in 36 hr, the detection limit can reach a single seeding cell (Figure 16-B). Therefore, in comparison to the culture-based biochemical tests for *F. nucleatum* detection, our RFD-FN1 system is not only simple-to-use but also less time consuming.

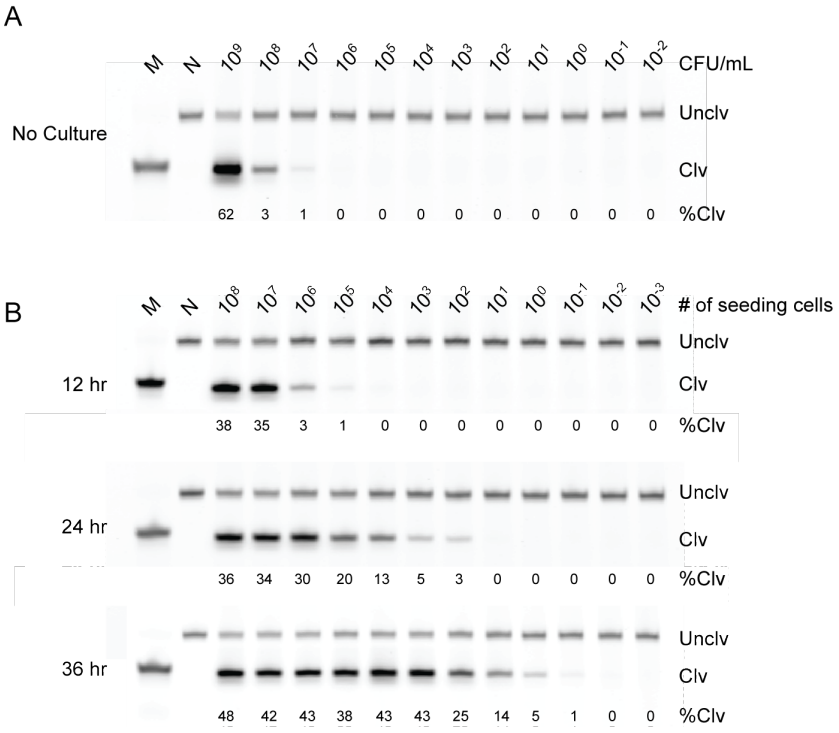

Figure 16. *Detection sensitivity of RFD-FN1.* (A) Detection limit in CFU/mL of DT4 in the presence of CIM-FN without additional cultivation. (B) Detection limit in the number of seeding cells of DT4 in the presence of CIM-FN with additional culturing time of 12, 24 and 36 hr.

3.2.7 Detection of *F. nucleatum* in fecal sample

As discussed in the introductory chapter, stool samples represent an attractive matrix for the development of a noninvasive detection method for CRC. Therefore, the functionality of RFD-FN1 in the presence stool sample was tested. Roughly 10^8 *F. nucleatum* cells were spiked into 50 mg of stool sample from a healthy donor. The mixture was homogenized and filtered through a 40 μm cell strainer to remove solid particles. The filtrate was spun down and the resulting cell pellet was processed to collect CIM as previously described. The resulting CIM was used to induce the activity of DT4. As shown in Figure 17, the presence of fecal sample completely abolished the activity of DT4. As a positive control, we tested the effect of the extraction process on the cleavage activity in the absence of stool samples. A reduction of cleavage activity likely due to the loss of cells during extraction was observed. However, this result suggests that the extraction method does not interfere with the activity of DT4 to a degree that it will completely abolish its activity. Therefore, the complete loss of activity is likely due to the presence of fecal samples. It is possible that there are inhibitory factors in the stool that interfere with the RFD-FN1 assay and a detailed investigation will be followed.

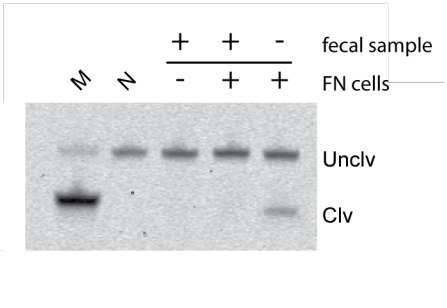

Figure 17. *Detection of F. nulceatum in the presence human stool sample.* The activity of RFD-FN1 DT4 was tested in the following reactions. N: negative control with reaction buffer alone. FN cells: with (+) or without (-) 10^8 *F. nucleatum* cells added to a test sample. Fecal sample: with (+) or without (-) 50 mg of fecal sample from a healthy donor.

3.2.8 Project summary

In summary, through in vitro selection using the crude extracellular matrix of *F. nucleatum*, we have isolated an RNA-cleaving DNAzyme, named RFD-FN1, that is activated by a protein marker. Interestingly, the protein target of RFD-FN1 is resistant to heat denaturation where 50% protein remains functional after being heated at 90ºC for 20 min. This is a highly valuable trait for direct human biological sample detection because these samples contain significant levels of nucleases that can be heat inactivated. Through detailed sequence analysis, the structure of RFD-FN1 in association with its substrate was presented. It was found that the catalytic core of RFD-FN1 is highly rich in A and T, which is rarely reported for similar classes of DNAzymes in literature. RFD-FN1 is also highly specific for *F. nucleatum* and it has a limit of detection of 10^7 CFU/mL without culture and one cell when cultured for 36 hr. Even though the utility of RFD-FN1 in the presence of human fecal sample is still under optimization, the discovery of this novel molecular probe for *F. nucleatum* presents opportunities for developing simple biosensors for this important pathogen, and the potential for CRC detection.

Chapter 4. Conclusion and Outlook

In this thesis, I have examined two strategies for developing a DNAzyme-based detection method for colorectal cancer. In the first project, crude human fecal samples were used as in vitro selection target for isolating CRC-specific RNA-cleaving DNAzymes. Results from this study revealed that high levels of nucleases are present in human fecal samples, which significantly interfere with the selection process. These findings suggest that an alternative selection strategy is needed to overcome nucleolytic degradation of DNA sequences. In the second project, I conducted an in vitro selection experiment using crude extracellular matrix of a gram-negative bacterium, *Fusobacterium nucleatum*, due to its prevalence in colorectal cancer patients. Through this selection, a highly specific RNA-cleaving DNAzyme, named RFD-FN1, was isolated. It was found that RFD-FN1 is activated by a heat resistant protein target inside *F. nucleatum* cells and that it can detect 10^7 CFU/mL without culture and one cell when cultured for 36 hr. Even though the utility of RFD-FN1 in the presence of human fecal sample is still under optimization, the discovery of this novel molecular probe for *F. nucleatum* presents exciting opportunities for developing fast and simple detection for this pathogen. Due to the implications of *F. nucleatum* in CRC pathogenesis, RFD-FN1 may be used as a starting point for further development of a simple screening method for CRC. Given the enhanced survival rate of early CRC detection, this probe can prove to be an invaluable tool in CRC disease management.

References

1. Ferlay, J.; Shin, H. R.; Bray, F.; Forman, D.; Mathers, C.; Parkin, D. M. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer.* **2010**, *127*, 2893-2917.
2. Fearon, E. R.; Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell.* **1990**, *61*, 759-767.
3. Hanahan, D.; Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell.* **2011**, *144*, 646-674.
4. Qin, J.; Li, R.; Raes, J.; Arumugam, M.; Burgdorf, K. S.; Manichanh, C.; Nielsen, T.; Pons, N.; Levenez, F.; Yamada, T.; *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature.* **2010**, *464*, 59-65.
5. Arthur, J. C.; Jobin, C. The struggle within: microbial influences on colorectal cancer. *Inflamm Bowel Dis.* **2011**, *17*, 396-409.
6. Dove, W. F.; Clipson, L.; Gould, K. A.; Luongo, C.; Marshall, D. J.; Moser, A. R.; Newton, M. A.; Jacoby, R. F. Intestinal neoplasia in the ApcMin mouse: independence from the microbial and natural killer (beige locus) status. *Cancer Res.* **1997**, *57*, 812-814.
7. Uronis, J. M.; Muhlbauer, M.; Herfarth, H. H.; Rubinas, T. C.; Jones, G. S.; Jobin, C. Modulation of the intestinal microbiota alters colitis-associated colorectal cancer susceptibility. *PLoS One.* **2009**, *4*, e6026.
8. Li, Y.; Kundu, P.; Seow, S. W.; de Matos, C. T.; Aronsson, L.; Chin, K. C.; Karre, K.; Pettersson, S.; Greicius, G. Gut microbiota accelerate tumor growth via c-jun and STAT3 phosphorylation in APCMin/+ mice. *Carcinogenesis.* **2012**, *33*, 1231-1238.
9. Gold, J. S.; Bayar, S.; Salem, R. R. Association of Streptococcus bovis bacteremia with colonic neoplasia and extracolonic malignancy. *Arch Surg.* **2004**, *139*, 760-765.
10. Zumkeller, N.; Brenner, H.; Zwahlen, M.; Rothenbacher, D. Helicobacter pylori infection and colorectal cancer risk: a meta-analysis. *Helicobacter.* **2006**, *11*, 75-80.

11. Wu, S.; Morin, P. J.; Maouyo, D.; Sears, C. L. Bacteroides fragilis enterotoxin induces c-Myc expression and cellular proliferation. *Gastroenterology.* **2003**, *124*, 392-400.
12. Maddocks, O. D.; Short, A. J.; Donnenberg, M. S.; Bader, S.; Harrison, D. J. Attaching and effacing Escherichia coli downregulate DNA mismatch repair protein in vitro and are associated with colorectal adenocarcinomas in humans. *PLoS One.* **2009**, *4*, e5517.
13. Eckburg, P. B.; Bik, E. M.; Bernstein, C. N.; Purdom, E.; Dethlefsen, L.; Sargent, M.; Gill, S. R.; Nelson, K. E.; Relman, D. A. Diversity of the human intestinal microbial flora. *Science.* **2005**, *308*, 1635-1638.
14. Turnbaugh, P. J.; Hamady, M.; Yatsunenko, T.; Cantarel, B. L.; Duncan, A.; Ley, R. E.; Sogin, M. L.; Jones, W. J.; Roe, B. A.; Affourtit, J. P.; Egholm, M.; Henrissat, B.; Heath, A. C.; Knight, R.; Gordon, J. I. A core gut microbiome in obese and lean twins. *Nature.* **2009**, *457*, 480-484.
15. Marchesi, J. R.; Dutilh, B. E.; Hall, N.; Peters, W. H.; Roelofs, R.; Boleij, A.; Tjalsma, H. Towards the human colorectal cancer microbiome. *PLoS One.* **2011**, *6*, e20447.
16. Kostic, A. D.; Gevers, D.; Pedamallu, C. S.; Michaud, M.; Duke, F.; Earl, A. M.; Ojesina, A. I.; Jung, J.; Bass, A. J.; Tabernero, J.; *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res.* **2012**, *22*, 292-298.
17. Castellarin, M.; Warren, R. L.; Freeman, J. D.; Dreolini, L.; Krzywinski, M.; Strauss, J.; Barnes, R.; Watson, P.; Allen-Vercoe, E.; Moore, R. A.; Holt, R. A. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res.* **2012**, *22*, 299-306.
18. Chen, W.; Liu, F.; Ling, Z.; Tong, X.; Xiang, C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS One.* **2012**, *7*, e39743.

19. Martinez, I.; Wallace, G.; Zhang, C.; Legge, R.; Benson, A. K.; Carr, T. P.; Moriyama, E. N.; Walter, J. Diet-induced metabolic improvements in a hamster model of hypercholesterolemia are strongly linked to alterations of the gut microbiota. *Appl Environ Microbiol.* **2009**, *75*, 4175-4184.
20. Turnbaugh, P. J.; Ridaura, V. K.; Faith, J. J.; Rey, F. E.; Knight, R.; Gordon, J. I. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med.* **2009**, *1*, 6ra14.
21. Pryde, S. E.; Duncan, S. H.; Hold, G. L.; Stewart, C. S.; Flint, H. J. The microbiology of butyrate formation in the human colon. *FEMS Microbiol Lett.* **2002**, *217*, 133-139.
22. Rose, D. J.; DeMeo, M. T.; Keshavarzian, A.; Hamaker, B. R. Influence of dietary fiber on inflammatory bowel disease and colon cancer: importance of fermentation pattern. *Nutr Rev.* **2007**, *65*, 51-62.
23. Wang, T.; Cai, G.; Qiu, Y.; Fei, N.; Zhang, M.; Pang, X.; Jia, W.; Cai, S.; Zhao, L. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J.* **2012**, *6*, 320-329.
24. Rubinstein, M. R.; Wang, X.; Liu, W.; Hao, Y.; Cai, G.; Han, Y. W. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. *Cell Host Microbe.* **2013**, *14*, 195-206.
25. Han, Y. W.; Shi, W.; Huang, G. T.; Kinder Haake, S.; Park, N. H.; Kuramitsu, H.; Genco, R. J. Interactions between periodontal bacteria and human oral epithelial cells: Fusobacterium nucleatum adheres to and invades epithelial cells. *Infect Immun.* **2000**, *68*, 3140-3146.
26. McCoy, A. N.; Araujo-Perez, F.; Azcarate-Peril, A.; Yeh, J. J.; Sandler, R. S.; Keku, T. O. Fusobacterium is associated with colorectal adenomas. *PLoS One.* **2013**, *8*, e53653.
27. Warren, R. L.; Freeman, D. J.; Pleasance, S.; Watson, P.; Moore, R. A.; Cochrane, K.; Allen-Vercoe, E.; Holt, R. A. Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome.* **2013**, *1*, 16.

28. Tahara, T.; Yamamoto, E.; Suzuki, H.; Maruyama, R.; Chung, W.; Garriga, J.; Jelinek, J.; Yamano, H. O.; Sugai, T.; An, B.; *et al.* Fusobacterium in colonic flora and molecular features of colorectal carcinoma. *Cancer Res.* **2014**, *74*, 1311-1318.
29. Ahn, J.; Sinha, R.; Pei, Z.; Dominianni, C.; Wu, J.; Shi, J.; Goedert, J. J.; Hayes, R. B.; Yang, L. Human gut microbiome and risk for colorectal cancer. *J Natl Cancer Inst.* **2013**, *105*, 1907-1911.
30. Kostic, A. D.; Chun, E.; Robertson, L.; Glickman, J. N.; Gallini, C. A.; Michaud, M.; Clancy, T. E.; Chung, D. C.; Lochhead, P.; Hold, G. L.; *et al.* Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe.* **2013**, *14*, 207-215.
31. Strauss, J.; Kaplan, G. G.; Beck, P. L.; Rioux, K.; Panaccione, R.; Devinney, R.; Lynch, T.; Allen-Vercoe, E. Invasive potential of gut mucosa-derived Fusobacterium nucleatum positively correlates with IBD status of the host. *Inflamm Bowel Dis.* **2011**, *17*, 1971-1978.
32. Zeller, G.; Tap, J.; Voigt, A. Y.; Sunagawa, S.; Kultima, J. R.; Costea, P. I.; Amiot, A.; Bohm, J.; Brunetti, F.; Habermann, N.; *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol.* **2014**, *10*, 766.
33. Zackular, J. P.; Rogers, M. A.; Ruffin, M. T. T.; Schloss, P. D. The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prev Res.* **2014**, *7*, 1112-1121.
34. Siegel, R.; Naishadham, D.; Jermal, A. Cancer statistics, 2012. *CA Cancer J Clin.* **2012**, *62*, 10-29.
35. Canadian Task Force on Preventive Health Care, Colorectal cancer screening. Recommendation statement from the Canadian Task Force on Preventive Health Care. *CMAJ.* **2001**, 165, 206-208.
36. Siegel, R.; DeSantis, C.; Virgo, K.; Stein, K.; Mariotto, A.; Smith, T.; Cooper, D.; Gansler, T.; Lerro, C.; Fedewa, S.; Lin, C.; Leach, C.; Spillers Cannady, R.; Cho, H.; Scoppa, S.; Hachey, M.; Kirch, R.; Jemal, A.; Ward, E. Cancer treatment and survivorship statistics, 2012. *CA Cancer J Clin.* **2012**, 62, 220-241.

37. Citarda, F.; Tomaselli, G.; Capocaccia, R.; Barcherini, S.; Crespi, M.; Group, T. I. Efficacy in standard clinical practice of colonoscopic polypectomy in reducing colorectal cancer incidence. *Gut.* **2001**, *48*, 812-815.
38. Rutter, C. M.; Johnson, E.; Miglioretti, D. L.; Mandelson, M. T.; Inadomi, J.; Buist, D. S. Adverse events after screening and follow-up colonoscopy. *Cancer Causes Control.* **2012**, *23*, 289–96.
39. Ahlquist, D. A.; Shuber, A. P. Stool screening for colorectal cancer: Evolution from occult blood to molecular markers. *Clinica Chimica Acta.* **2002**, *315*, 157-68.
40. Walsh, J. M.; Terdiman, J. P. Colorectal cancer screening: scientific review. *JAMA*. **2003**, 12,1288-96.
41. Ouyang, D. L.; Chen, J. J.; Getzenberg, R. H.; Schoen, R. E. Noninvasive testing for colorectal cancer: a review. *Am J Gastroenterol*. **2005**, 100, 1393-403.
42. Hermann, T.; Patel, D. J. Adaptive recognition by nucleic acid aptamers, *Science*. **2000**, 287, 820-825.
43. Ponce-Salvatierra, A.; Wawrzyniak-Turek, K.; Steuerwald, U.; Höbartner, C.; Pena, V. Crystal structure of a DNA catalyst, *Nature*. **2016**, 529, 231-234.
44. Robertson, D. L. and Joyce, G. F. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*. **1990**, 344, 467–468.
45. Ellington, A. D. and Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature*. **1990**, 346, 818–822.
46. Tuerk, C. and Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*. **1990**, 249, 505–510.
47. Liu, J.; Cao, Z.; Lu, Y. Functional nucleic acid sensors, *Chem. Rev.* **2009**, 109, 1948-1998.
48. Schlosser, K.; Li, Y. Biologically inspired synthetic enzymes made from DNA. *Chem. Biol.* **2009**, 16, 311-322.
49. Navani, N. K.; Li, Y. Nucleic acid aptamers and enzymes as sensors. *Curr. Opin. Chem. Biol.* **2006**, 10, 272-281.

50. Joyce, G. F. Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem*. **2004**, 73, 791-836.
51. Wilson, D. S and Szostak, J. W. In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* **1999**, 68, 611-47.
52. Mei, S. H.; Liu, Z.; Brennan, J. D.; Li, Y. An efficient RNA-cleaving DNA enzyme that synchronizes catalysis with fluorescence signaling. *J Am Chem Soc.* **2003**, *125*, 412-420.
53. Li, Y. and Breaker, R. R. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2′-hydroxyl group. *J. Am. Chem. Soc*. **1999**, 121, 5364–5372.
54. Forster, A. C.; Symons, R. H. Self-cleavage of plus and minus RNA of a virusoid and a structural model for the active sites. *Cell*. **1987**, 49, 211-220.
55. Hampel, A.; Tritz, R. RNA catalytic properties of the minimum (-)sTRSV sequence. *Biochemistry*. **1989**, 28, 4929-4933.
56. Kuo, M. Y.; Sharmeen, L.; Dinter-Gottlieb, G.; Taylor, J. Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *J. Virol*. **1988**, 62, 4439- 4444.
57. Saville, B. J.; Collins, R. A. A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell*. **1990**, 61, 685-696.
58. Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*. **2004**, 428, 281-286.
59. Silverman, S. K. In vitro selection, characterization, and application of deoxyribozymes that cleave RNA. *Nucleic Acids Res*. **2005**, 33, 6151-6163.
60. Schlosser, K. and Li, Y. A versatile endoribonuclease mimic made of DNA: characteristics and applications of the 8-17 RNA-cleaving DNAzyme. *ChemBioChem*. **2010**, 11, 866-879.
61. Breaker, R. R. and Joyce, G. F. A DNA enzyme that cleaves RNA. *Chem. Biol*. **1994**, 1, 223-229.

62. Santoro, S. W. and Joyce, G. F. A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. USA*, **1997**, 94, 4262–4266.
63. Ali, M. M.; Aguirre, S. D.; Lazim, H.; Li, Y. Fluorogenic DNAzyme probes as bacterial indicators. *Angew. Chem. Int. Ed*. **2011**, 50, 3751-3754.
64. Tram, K.; Kanda, P.; Salena, B. J.; Huan, S.; Li, Y. Translating bacterial detection by DNAzymes into a litmus test. *Angew. Chem. Int. Ed*. **2014**, 53, 12799-12802.
65. He, S.; Qu, L.; Shen, Z.; Tan, Y.; Zeng, M.; Liu, F.; Jiang, Y.; Li, Y. Highly specific recognition of breast tumors by an RNA-cleaving fluorogenic DNAzyme probe. *Anal. Chem*. **2015**, 87, 569-577.
66. Shen, Z.; Wu, Z.; Chang, D.; Zhang, W.; Tram, K.; Lee, C.; Kim, P.; Salena, B. J.; Li, Y. A catalytic DNA activated by a specific strain of bacterial pathogen. *Angew. Chem. Int. Ed.* **2016**, 55, 2431-4.
67. Liu, Z.; Mei, S. H. J.; Brennan, J. D.; Li, Y. Assemblage of signaling DNA enzymes with intriguing metal specificity and pH dependence. *J. Am. Chem. Soc*. **2003**, 125, 7539-7545.
68. Achenbach, J. C.; Nutiu, R.; Li, Y. Structure-switching allosteric deoxyribozymes. *Anal. Chim. Acta*. **2005**, 534, 41-51.
69. Shen, Y.; Chiuman, W.; Brennan, J. D.; Li, Y. Catalytic and signaling properties of trans-acting pH6DZ1, an RNA-cleaving and fluorescence-signaling deoxyribozymes with a four-way junction structure. *ChemBioChem*. **2006**, 7, 1343-1348.
70. Chiuman, W.; Li, Y. Simple fluorescent sensors engineered with catalytic DNA 'MgZ' based on a non-classic allosteric design. *PLoS One*. **2007**, 2, e1224.
71. Li, Y. Advancements in using reporter DNAzymes for identifying pathogenic bacteria at speed and with convenience. *Future Microbiol*. **2011**, 6, 973-976.
72. Bolstad, A. I.; Jensen, H. B.; Bakken, V. Taxonomy, biology, and periodontal aspects of Fusobacterium nucleatum. *Clin Microbiol Rev.* **1996**, 9, 55-71.
73. Jervøe-Storm, P. M.; Koltzscher, M.; Falk, W.; Dörfler, A.; Jepsen, S. Comparison of culture and real-time PCR for detection and quantification of five putative

periodontopathogenic bacteria in subgingival plaque samples. *J Clin Periodontol.* **2005**, 32(7), 778-83.

74. Boutaga, K.; van Winkelhoff, A. J.; Vandenbroucke-Grauls, C. M.; Savelkoul, P. H. Periodontal pathogens: a quantitative comparison of anaerobic culture and real-time PCR. *FEMS Immunol Med Microbiol.* **2005**, 45(2), 191-9.
75. Topcuoglu, N.; Paltura, C.; Kulekci, M.; Ustek, D.; Kulekci, G. Real-time polymerase chain reaction versus conventional PCR: a comparison between two methods for the detection of *Fusobacterium nucleatum* in saliva, nasopharyngeal secretion and middle ear effusion samples. *Biotechnol. Biotechnol. Equip.* **2013**, 27, 3825–3828.
76. Schmieder, R. and Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* **2011**, 27, 863-4.
77. Li, W. and Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. **2006**, 22, 1658-9.
78. Yang, J.-H.; Usman, N.; Chartrand, P.; Cedergren, R. J. Minimum ribonucleotide requirement for catalysis by the RNA hammerhead domain. *Biochemistry*. **1992**, *31*, 5005–5009
79. Ortigao, J. F. R.; Rosch, H.; Selter, H.; Frohlich, A.; Lorenz, A.; Montenarh, M.; Seliger, H. Antisense effect of oligodeoxynucleotides with inverted terminal internucleotidic linkages: a minimal modification protecting against nucleolytic degradation. *Antisense Res. Dev.* **1992**, *2*, 129-146.
80. Schubert, S.; Gul, D. C.; Grunert, H. -P.; Zeichhardt, H.; Erdmann, V. A.; Kurreck, J. RNA cleaving '10-23' DNAzymes with enhanced stability and activity. *Nucleic Acids Res.* **2003**, *31*, 5982-5992.
81. Lan, T. and Lu, Y. Metal ion-dependent DNAzymes and their applications as biosensors. *Met Ions Life Sci*. **2012**, 10, 217-48.
82. Aguirre, S. D.; Ali, M. M.; Salena, B. J.; Li, Y. A sensitive DNA enzyme-based fluorescent assay for bacterial detection. *Biomolecules*. **2013**, 20, 563-77.