Investigation and characterization of functional nucleic acids in whole human serum for the detection of biomarkers towards diagnostic application

By

Ioana Cozma, MD BScH

A Thesis Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2018) McMaster University

(Biochemistry and Biomedical Sciences) Hamilton, Ontario

TITLE: Investigation and characterization of functional nucleic acids in whole human serum for the detection of biomarkers towards diagnostic application

AUTHOR: Ioana Cozma, MD (McMaster University), BScH (Queen's University)

SUPERVISOR: Dr. Yingfu Li

NUMBER OF PAGES: xvii, 211

Abstract

Steady advancements in diagnostics over the past century have propelled the world of medicine into the more advanced era of preventative medicine, an era with a resoundingly clear message: early detection can save lives. For patients who suffer from either pancreatic cancer or malignant hyperthermia susceptibility, early or preoperative diagnosis, respectively can save lives and minimize morbidity and mortality, in addition to offering cost-savings to hospitals and healthcare systems. Fortunately, significant progress have been made in the fields of metabolomics and biomarker identification. Given the benefits carried by serum biomarkers as targets of screening and diagnostic tool development, we applied functional nucleic acid technology and *in vitro* selection directly in whole human serum to search for disease-specific biomarkers and associated detection probes without *a priori* knowledge of the biomarkers pursued. This endeavour simultaneously serves as a proof-of-concept study to establish whether *in vitro* selection can be successfully performed in human serum.

We specifically focused on the derivation of RNA-cleaving DNAzymes (RCD) through *in vitro* selection, or SELEX (systemic evolution of ligands through exponential exposure). DNAzymes constructed with a fluorogenic signalling molecule were incubated with human serum with the goal of identification of a functional nucleic acid probe capable of detecting the presence of a disease-specific biomarker. Two independent protocols have been designed and executed for the identification of DNAzyme sequences capable of detecting pancreatic cancer and malignant hyperthermia susceptibility, respectively.

The first exploration was performed in serum obtained from cancer patients, with the goal of identifying DNAzymes capable of distinguishing pancreatic cancer from other cancer types. To do so, we employed *in vitro* selection, Next-Generation Sequencing, and bioinformatic analysis. We successfully demonstrated the feasibility of performing *in vitro* selection with DNAzymes in human serum, evidenced by distinct round-to-round enrichment of a DNA library towards the identification of DNAzymes capable of detecting pancreatic cancer. Additionally, we isolated two DNAzymes capable of distinguishing pancreatic cancer serum from healthy patient serum in fresh collected serum samples.

Based on the positive results gathered in the pancreatic cancer *in vitro* selection project, we subsequently endeavoured to replicate the demonstrated feasibility of performing *in vitro* selection in human serum. By selecting malignant hyperthermia as the pathology investigated, we simultaneously sought to diversify the scope of DNAzyme detection by establishing whether successful DNAzyme selection can be achieved in a non-acute disease state. Thus, the second exploration was performed in serum obtained from patients who underwent evaluation for malignant hyperthermia susceptibility using the gold-standard caffeine-halothane contracture test. The goal of this project rested on the identification of DNAzymes capable of distinguishing malignant hyperthermia susceptibility in serum and approximating the performance of the gold standard test. We successfully isolated four DNAzyme candidates which demonstrated clinically relevant thresholds of sensitivity and specificity following thorough sensitivity and specificity analysis. In doing so, we once again demonstrated the ability to perform *in vitro* selection in human serum.

Given the complexity of molecular interactions observed over the course of two *in vitro* selection protocols in human serum, it became clear that distinguishing meaningful target-mediated interactions from non-specific interactions would require advanced bioinformatic analysis. Consequently, using principles of computational biology, we performed a deep exploration of Next-Generation Sequencing results obtained from sequencing our recovered DNA libraries to extract additional data that would inform on the next required steps required to identify a DNAzyme specific for the pathology pursued. In doing so, we identified a two-step method to evaluate the progress of the *in vitro* selection protocol undertaken, and offered a systematic approach for choosing candidate sequences to undergo further testing based on promising performance *in silico*. Using this approach, we successfully identified a DNAzyme sequence capable of acting as a general cancer detection probe, with promising potential for diagnostic application.

Ultimately, this thesis serves as a feasibility study of a novel approach to both *in vitro* selection and biomarker identification technique by combining the latest nanotechnology techniques with clinical data and real patient serum samples, and advanced computational biology tools. Despite the inability to identify a highly sensitive and specific DNAzyme capable of advancing towards biosensor construction, several important strides and lessons have been acknowledged, establishing the feasibility of performing *in vitro* selection in human serum, and outlining strategies for addressing and anticipating challenges with this technique. The hope is for this work to inspire and inform future efforts to apply functional nucleic acid technology to solve current gaps in both the diagnostic and therapeutic branches of medicine, and with the help of computational biology continue to bridge the gap between basic science and clinical medicine.

Acknowledgements

I would like to thank my supervisor Dr. Yingfu Li for his constant guidance and mentorship over the past five years. I am incredibly lucky to have had the opportunity to work under the supervision of an acclaimed researcher, who fosters such an inclusive and supportive environment for growth and curiosity. I am furthermore incredibly grateful for his trust in me to pursue high-risk scientific endeavours, and for shaping me into the clinician scientist I am today. I would also like to thank my committee members Dr. Bruno Salena, Dr. Sheila Singh, and Dr. Sheila Riazi for their continued support, guidance, and encouragement. You are all inspirations to me in your commitment to research balanced with your passion for clinical work. I hope to one day be able to emulate the qualities you emanate.

I would also like to thank Jim Gu, without whom this work would not be possible. In particular, I would like to acknowledge Jim's dutiful work generating the bioinformatic analysis and taking the time to teach me the techniques employed, in addition to patiently training me on the basic science techniques required to conduct this thesis work. Thank you for also taking the time to answer my many questions throughout the past 5 years and for helping me grow into a successful scientist.

I additionally need to thank the members of the Li Lab for making me look forward to coming to the lab every day, especially the *DNAzyme Dames* (Erin, Sandy, Devon, Meghan) for fostering an environment of collaboration, empowerment, and friendship. Furthermore, I would like to express my gratitude to my Undergrad Army for your hard work and commitment to helping me complete my experiments in a short timeline.

I would like to express my thanks to the McMaster Department of Anesthesia, for taking a chance on me and supporting me through my academic and research pursuits. I would also like to express my appreciation to Dr. Zena Samaan, the program director of the Clinician Investigator Program for the continued encouragement and guidance throughout a tumultuous five years.

Finally, I would like to thank my husband, Stefan for his never-ending love and support of every ambition I pursue. There are no words to describe my gratitude for every dinner brought to the lab, every gel plate washed, every pep-talk, every late-night pick-up when I was too tired to drive home, and for always empowering me to be the best version of myself. Thank you for always being willing to carefully listen and help me bounce ideas and theories.

I would additionally like to express my gratitude to my parents for their boundless love, support, encouragement, and for inspiring a love of knowledge and service to my community since childhood. Thank you for fuelling my ambition and always championing every academic and non-academic pursuit. I am forever grateful and inspired by your sacrifices to give me a better life, and I hope I can continue to make you proud.

To Stefan, my everything

To Sorin and Concetta, my anchors and pillars of support

To Luna, my sunshine

Table of Contents

Declaration of Academic Achievement

I have taken the primary role in the design of the projects conducted in this thesis, in the analysis and interpretation of the data, and in the writing of the manuscripts, with the exception of the bioinformatic analysis implementation, which was conducted by Jim Gu. All experimental work conducted in this thesis work was performed by me.

LIST OF FIGURES

Chapter 3

Chapter 4

LIST OF TABLES

LIST OF ABBREVIATIONS

ADH	Alcohol Dehydrogenase
ALPPL-2	Alkaline Phosphatase Placental-Like 2
ALS	Amyotrophic Lateral Sclerosis
APE1	Human Ap-Endonuclease 1
ATP	Adenosine Triphosphate
CA 19-9	Carbohydrate 19-9
CA 195	Cancer Antigen 195
CA 242	Cancer Antigen 242
CA 50	Cancer Antigen 50
CA 72-4	Cancer Antigen 72-4
CA-125	Cancer Antigen 125
CAM 17.1	Cancer Biomarker
CD-HIT-EST	Clustering Algorithm
CEA	Carcinoembryonic Antigen
CEMIP	Cell Migration-Inducing And Hyaluronan-Binding Protein
CHCT	Caffeine-Halothane Contracture Test
CLEC3B	Tetranectin
COVID-19	Coronavirus Disease 2019
CRC	Colorectal Cancer
CT	Computer Tomography
CTCR	Cleavage Time Course Reaction
CXCL9	Chemokine Ligand 9
DNA	Deoxyribonucleic Acid
DREME	Discriminative Regular Expression Motif Elicitation
DUPAN-2	Duke Pancreatic Mono-Clonal Antigen Type 2
EDTA	Ethylenediaminetetraacetic Acid
EGFR	Estimated Glomerular Filtration Rate
FBS	Fetal Bovine Serum
FNA	Functional Nucleic Acid
GI	Gastrointestinal
GR-5	Gr-5 DNAzyme
HDLBP	High Density Lipoprotein Binding Protein
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
HGF	Hepatocyte Growth Factor
HIREB	Hamilton Integrated Research Ethics Board
IDT	Integrated DNA Technologies
IL	Interleukin

LDH	Lactate Dehydrogenase
L1CAM	L1 Cell Adhesion Molecule
MEME	Multiple Expectation Maximizations For Motif Elicitation
MH	Malignant Hyperthermia
MHIU	Malignant Hyperthermia Investigation Unit
MHN	Malignant Hyperthermia Negative
MHS	Malignant Hyperthermia Susceptible
MIC-1	Macrophage Inhibitory Cytokine
MRI	Magnetic Resonance Imaging
MUC	Mucin
MUSCLE	Bioinformatic Tool
NGS	Next-Generation Sequencing
NH	Human Serum
OTB	Ontario Tumour Bank
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDA	Pancreatic Ductal Adenocarcinoma
PET	Positron Emission Tomography
PNET	Pancreatic Neuroendocrine Tumour
PNK	T4 Polynucleotide Kinase
POA	Pancreatic Oncofetal Antigen
PSA	Prostate-Specific Antigen
PTK7	Tyrosine Kinase 7
QB	Quenching Buffer
RCD	RNA-Cleaving DNAzymes
RFD	RNA-Cleaving Fluorogenic DNAzyme
RNA	Ribonucleic Acid
SB	Selection Buffer
SDS-PAGE	Sodium Dodecyl-Sulfate Polyacrylamide Gel Electrophoresis
SELEX	Systemic Evolution Of Ligands Through Exponential Exposure
SPAN-1	Cancer Biomarker
STIP1	Stress-Induced Phosphoprotein 1
TATI	Tumor-Associated Trypsin Inhibitor
TPS	Tissue Polypeptide Specific Antigen
VEGF	Vascular Endothelial Growth Factor
YKL-40	Chitinase-3-Like Protein 1

Chapter 1. Introduction

1.1 Pancreatic Cancer

Pancreatic cancer has been widely recognized as the most lethal type of common cancer, with a dismal 5 year survival rate of only 12%.[1,2] Ranking as the 4th leading cause of cancer related deaths, its incidence largely matches its mortality.[2] In 2020 the global incidence of pancreatic cancer was 500,000 with a calculated incidence rate of 4.9 per 100,000, while the mortality of pancreatic cancer was 470,000 with a calculated mortality rate of 4.5 per 100,000. In Canada, estimates show that in 2022, 6,900 new cases of pancreatic cancer will be diagnosed, and 5,700 Canadians will lose their battle with pancreatic cancer.[3] The poor prognosis of pancreatic cancer is attributed to 3 factors: late diagnosis, aggressive nature of disease, and poor response to treatment.[4] Ultimately however, the driving force behind the morbidity and mortality associated with pancreatic cancer is late diagnosis,[2] as the vast majority of patients are only diagnosed following the onset of symptoms.

The presence of symptoms, especially obstructive symptoms like abdominal pain and jaundice, is already an indicator of advanced disease[2], which limits treatment options and too often eliminates any feasibility of cure.[4] Currently, the only treatment with curative intent is surgical resection, followed by chemotherapy or chemoradiation[2]. At the time of diagnosis, only 20% of patients have resectable disease, and of these only half go on to have successful resection.[4] Despite successful resection and adjuvant therapy, the 5 year survival still only ranges from 3-44%.[4] The remaining patients with borderline resectable, non-resectable disease (locally advanced), or metastatic disease go on to have chemotherapy, chemoradiation, or palliation.[2] Virtually all pancreatic cancer

patients require palliation, whether it is from upfront aggressive and/or metastatic disease or post-treatment recurrence.[2]

Pancreatic cancer can be divided into two categories based on the cell of origin.[1] As such, pancreatic cancer tumours can be exocrine or neuroendocrine.[1] The vast majority of pancreatic cancer tumours are exocrine tumours, with pancreatic ductal adenocarcinoma (PDA) presiding as the most frequent type.[1] Neuroendocrine tumours (PNETs) amount to approximately 7% of pancreatic cancers. Unlike PDA which is characterized as a highly aggressive cancer responsible for the dire prognosis of pancreatic cancer secondary to treatment limitations, PNETs are slow-growing and much more treatable than PDA even when accompanied by metastatic deposits at initial diagnosis.[1]

Despite the grim prognosis and dire/alarming statistics, there is no current screening program in place for pancreatic cancer.[2] The conclusion from several meta-analyses and literature reviews is resoundingly clear: early detection can save lives and minimize morbidity and mortality.[2,5] Furthermore, future research efforts need to focus on advancing early detection and screening methods, along with instituting screening programs, especially for high-risk individuals, such as those with positive family history.[2] The current diagnostic tools for making a pancreatic cancer diagnosis are either too invasive, costly, or lack the sensitivity and specificity required for use in a screening program.[5] The gold standard diagnostic tool for pancreatic cancer is tissue biopsy, obtained via endoscopic ultrasound or endoscopic retrograde cholangiopancreatography.[6] Additional imaging modalities frequently used in the diagnostic workup of pancreatic cancer include computer tomography (CT), transabdominal ultrasound, magnetic resonance imaging

(MRI), and positron emission tomography (PET).[6] The biomarker described as having the highest degree of clinical usefulness in pancreatic cancer is CA 19-9.[5]

CA 19-9, or serum carbohydrate antigen 19-9, is characterized as a tumour-associated antigen circulating in the blood of patients with gastrointestinal (GI) cancers.[7] It was initially discovered in the 1970s and at the time carried significant promise for its potential use in early diagnosis and screening.[7] Though it is faithfully used in both the diagnostic workup and for tracking disease recurrence and response to chemotherapy and chemoradiation,[4,5] CA 19-9 is not sensitive nor specific enough to be debuted as a screening tool.[4] Clinically, only patients who are positive for Lewis antibody a or b are able to produce CA 19-9, which eliminates approximately 10% of the population, leading to false negative results.[5] Additionally, CA 19-9 can be falsely elevated in patients suffering from cholestasis or other benign hepatobiliary pathology[5], which impacts its specificity and false positive rate. CA 19-9 is often used in combination with CEA (carcinoembryonic antigen) in the diagnostic workup of pancreatic cancer.[2] Some studies have also shown added benefit to incorporating CA-125 (carbohydrate antigen 125), a biomarker used very frequently in the diagnosis of ovarian cancer.[7] CA-125 has been found to be clinically useful in diagnosing pancreatic cancer in patients who are not jaundiced, or who do not synthesize CA 19-9 in detectable quantities.[7] In fact, studies suggest that an assay combining any or all of CA 19-9, CEA, CA125, and additional biomarkers yet to be discovered may be the key to future screening programs and earlier diagnosis.[7]

Presently, several PDA serum biomarker contenders have been identified, and include: amylin, DUPAN-2, CA 242, CAM 17.1, TPS, CA 72-4, SPan-1, CA 50, CA 195, TATI, POA, and YKL-

40.[7] However, none of these were found to be superior to CA 19-9 in comparison assays.[7] Macrophage Inhibitory Cytokine 1 (MIC-1) was shown to be equivalent to CA 19-9 in separating patients with resectable disease from patients with benign periampullary pathology.[8] Additionally, combining MIC-1 and CA 19-9 has been proven to augment the efficacy of CA 19-9.[8] Additional validation studies however are required before clinical implementation of MIC-1.

1.2 Malignant Hyperthermia

Malignant hyperthermia (MH) is a highly lethal pharmacogenetic disorder transmitted in an autosomal dominant pattern, characterized by deranged Ca^{2+} homeostasis in skeletal muscle.[9] In the clinical setting, malignant hyperthermia reactions are triggered by exposure to volatile anesthetic gases such as sevoflurane, desflurane, and halothane, or by exposure to succinylcholine, a depolarizing muscle relaxant.[10] However, MH reactions can also be triggered outside the clinical setting by stresses caused by heat and vigorous exercise.[10]

The pathophysiology of MH can be traced back to mutations in *RYR1, CACNA1S* and *STAC3*[11] genes, whose products are involved in skeletal muscle Ca^{2+} regulation.[9] In most cases, a defect in the ryanodine receptor is responsible for the clinical manifestation of MH.[12] In affected individuals, exposure to triggering anesthetic agents results in excessive cytoplasmic calcium release in skeletal muscle, leading to a hypermetabolic response characterized by muscle rigidity and cellular breakdown (rhabdomyolysis).[13] Other features include excessive carbon dioxide production and oxygen consumption, severe hyperthermia, tachycardia, tachypnea, acidosis, and hyperkalemia.[13] Without treatment, a malignant hyperthermia reaction is lethal.[10] Consequently, MH it is recognized as one of the most dangerous perioperative anesthetic complications, with an

estimated event triggering incidence of 1/15,000 to 1/75,000, and affects patients of all ethnicities.[13–15] Calculating the prevalence of MH is challenging due to incomplete penetrance and variable expression of MH susceptible traits, but is estimated to occur at a rate of 1:200 to 1:3000.[13]

The major determinant of morbidity and mortality associated with an MH reaction is late diagnosis.[14] Unfortunately, the diagnosis of MH most commonly occurs only following symptom onset.[10] Treatment of an acute MH reaction consists of immediately terminating the triggering agent, administration of intravenous dantrolene, hyperventilation to normocapnia, and supportive measures for symptom control.[12] Such supportive measures include cooling protocols, antiarrhythmics, diuretics, and electrolyte corrections.[13] Dantrolene sodium is the only treatment available for an acute MH reaction, and its mechanism of action serves to block further release of calcium into the myoplasm, arresting the pathophysiologic response triggered.[10]

Presently there is a notable absence of accessible diagnostic tools to screen patients undergoing general anesthesia and detect MH susceptibility prior to a triggering event. Apart from clinical diagnosis following an MH triggered event, only 2 diagnostic tools are currently available: caffeine-halothane contracture test (CHCT) and genetic testing.[13]

The CHCT is an ex vivo diagnostic test performed on a muscle biopsy designed to simulate exposure to anesthesia and observe whether an MH reaction is triggered.[16] Consequently, the CHCT involves obtaining a fresh surgically excised muscle biopsy, which must be immediately dissected into strips and sutured to a force transducer.[17] Subsequently, the muscle strips are subjected to caffeine and halothane to observe whether a hypermetabolic response is triggered.

[17]Although the CHCT is 97-100% sensitive, it remains invasive (requiring a 7 x 3 x 1 cm biopsy of fresh muscle and 4-6 weeks recovery), costly, and requires extensive travel to one of only 30 specialized centers worldwide because the CHCT must be completed within 5 hours of muscle biopsy.[18] Due to these significant limitations, only about 4% of those who are suspected of being MH-susceptible undergo CHCT.[9]

Conversely MH genetic testing is less invasive and costly than the CHCT, but due to its low sensitivity of 50-60%, it remains a poor diagnostic test.[19] Initially, genetics showed tremendous potential, however extensive investigations, there are now more than 200 identified mutations in *RYR1, CACNA1S* and *STAC3,* the genes implicated in the development MH.[11] This is further complicated by the unknown functionality, genetic heterogeneity, and variable penetrance of these mutations.[20] Consequently, the utility of genetic testing continues to be very limited and there remains a great need for an easily accessible, reliable, inexpensive, and minimally invasive diagnostic tool. To date, there are no previously identified biomarkers of malignant hyperthermia. Subsequently, efforts aimed towards the identification and characterization of novel biomarkers of may be the key to the development of future minimally invasive diagnostic tools for screening and preoperative diagnosis.

1.3 Functional nucleic acids and *in vitro* selection

The serendipitous discovery of the ribozyme by both Thomas Cech and Sydney Altman in 1982 sent shockwaves through the fields of chemistry and molecular biology, bringing about two large paradigm shifts in our understanding of nucleic acids and enzymes.[21] The ribozyme, also known as an RNA enzyme, is a naturally occurring RNA molecule capable of performing catalytic functions.[22] Its discovery subsequently challenged the long-held notion that proteins are the only biological molecule capable of enzymatic activity.[22] Unsurprisingly, this discovery earned Cech and Altman the Nobel Prize in Chemistry in 1989.[21] The second paradigm shift caused by the discovery of the ribozyme was the notion of nucleic acids adopting/developing functional abilities beyond their conventional role in the storage, transfer, and expression of genetic material.[21] Consequently, a new category of nucleic acids – functional nucleic acids was introduced. Today, functional nucleic acids encompass a wide array of naturally occurring and synthetically derived oligonucleotide segments capable of ligand binding, catalysis, and regulatory functions.[23]

As alluded to, functional nucleic acids can be distinguished by their occurrence in nature, with ribozymes and riboswitches belonging to this category.[22] Ribozymes have been shown to have the capacity to catalyse a range of biochemical reactions, including the breaking and joining of phosphodiester bonds and peptide bonds.[22] They have further been demonstrated to retain their functional abilities even under hostile prebiotic conditions such as extremes of temperature, pH, pressure, and ultraviolet light, suggesting the development of nucleic acid functionality may be traced back to the primitive Earth.[22] Riboswitches, first discovered in the early 2000s, are RNA molecules capable of metabolite detection and associated regulation of gene expression.[24] Specifically, riboswitches have been shown to have adopted a cellular metabolite detection domain

(aptamer domain) coupled to an expression element, such that in the presence of certain metabolites which bind to the aptamer domain, a conformational change is induced in the expression element to enable or hinder gene expression.[21] Consequently, it appears riboswitches have been evolved to act as naturally derived biosensors capable of regulating metabolic processes in their immediate surroundings.[22] These properties, along with the discovery of ribozymes' retained functionality in prebiotic conditions have given rise to the "RNA world" hypothesis.[22] This hypothesis suggests that the evolution of life on Earth may have commenced with the spontaneous polymerization of existing ribonucleotides, slowly yielding chains of RNA oligonucleotides.[25] Through mutations and natural selection, these molecules evolved functionality to enable storage of genetic information and catalysis of metabolic reactions in their immediate surroundings, eventually including self-replication, all prior to the evolution of modern cells and the transition to DNA and proteins.[25]

Unsurprisingly, functional nucleic acid research skyrocketed following Cech and Altman's breakthrough. Consequently, a mere eight years after the seminal discovery of the naturally occurring ribozyme, the development of the first artificial functional nucleic acids was announced.[21] These synthetic functional nucleic acids were derived using a novel technique developed by Larry Gold in 1990 named *in vitro* selection, which revolutionized the field of molecular biology.[26] Through *in vitro* selection, otherwise known as SELEX (Systematic Evolution of Ligands by EXponential enrichment), scientists were now able to specifically isolate functional nucleic acids based on desired criteria or properties, such as ligand binding or catalysis[26]. This powerful systematic technique, illustrated in Figure 1, involves iterative rounds of exposure of a large synthetic pool of unique oligonucleotide sequences to a target or selection

pressure of interest in order to drive the evolution of a desired property.[26] Sequences demonstrating the desired binding or catalytic property are rewarded with amplification, while sequences that do not display the desired functionality are eliminated from the pool of oligonucleotides.[26] In effect, as the name indicates, the process of *in vitro* selection simulates natural selection in a test tube. Consequently, with every round of selection, the diversity of the pool of oligonucleotides is expected to shrink significantly compared to the starting pool, such that the product of *in vitro* selection yields a significantly diminished pool of oligonucleotides with expected functionality reflective of the selective pressures applied.[21] Furthermore, the most abundant sequences in the recovered pool following the completion of *in vitro* selection are expected correlate with more pronounced functionality.[27]

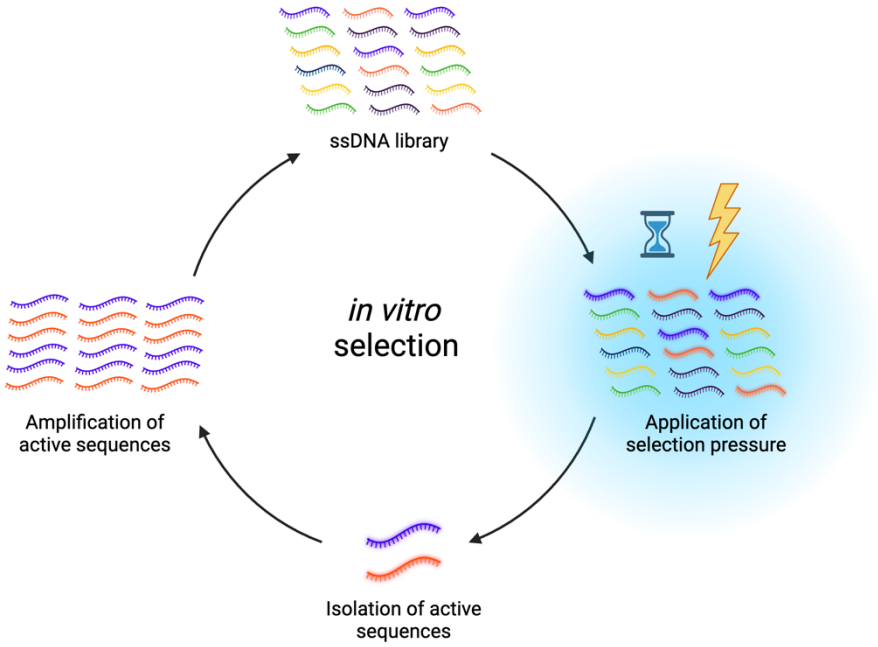


Figure 1. *in vitro* selection schematic for identification of DNA-based functional nucleic acids.

The advent of *in vitro* selection consequently introduced distinct new categories of synthetic functional nucleic acids, differentiated by their functional properties. Notable categories include aptamers and DNAzymes. Aptamers are the first synthetic functional nucleic acids isolated through *in vitro* selection, and are defined as single-stranded oligonucleotide sequences (RNA or DNA-based) distinguished by their ability to bind to a specific target with high affinity and fidelity.[27] The targets can range in size from metal ions to large proteins, demonstrating the versatility of aptamers.[28] DNAzymes or deoxyribozymes are DNA-based enzymes, defined as single-stranded oligonucleotide sequences capable of performing a catalytic function.[21] Unlike their RNA counterparts, DNAzymes are entirely artificial.[21]

The first DNAzyme identified was the GR-5 DNAzyme, constructed in 1994 by Gerald Joyce and Ronald Breaker.[29] GR-5 belongs to a subcategory of DNAzymes, specifically RNA-cleaving DNAzymes (RCDs), reflecting the ability of these functional nucleic acids to cleave an RNA substrate by catalyzing the transesterification reaction of its phosphodiester backbone.[29] The RNA substrate can be located within a separate RNA sequence or can be embedded within a DNA sequence.[29] Consequently, a DNAzyme catalyzing the cleavage of another oligonucleotide molecule is further classified as a *trans*-acting DNAzyme.[30] Conversely, the RNA substrate can be embedded within the sequence of the DNAzyme such that in the presence of a cofactor, the DNAzyme undergoes self-cleavage demonstrating the properties of a *cis*-acting DNAzyme.[30] As such, GR-5 can be classified as a *cis*-acting RNA-cleaving DNAzyme, which specifically catalyzes the Pb^{2+}-dependent cleavage of an RNA substrate embedded within its sequence.[29] Since the activity of GR-5 is dependent on the presence and concentration of Pb^{2+} in solution, it can

effectively act as a metallosensor, through the generation of a detectable cleavage fragment when exposed to specific concentrations of Pb^{2+}.[21]

Following the construction of the first RNA-cleaving DNAzyme, the research community quickly realized the potential RCDs hold for biosensing. In addition to their remarkable ability to identify femtomolar concentrations of a target,[21] DNAzymes are particularly advantageous due to their intrinsic functional stability, which is an important property when developing diagnostic assays or biosensors.[31] Additionally, their low cost of synthetic production is another appealing characteristic when considering downstream application.[21] This recognition soon saw the developments of the first RNA-cleaving Fluorogenic DNAzyme (RFD).[21] Fashioned with a fluorescence-based signalling molecule made up of a fluorophore and quencher pair, RFDs were engineered to emit a detectable fluorescent signal indicative of DNAzyme catalysis.[32] By embedding the RNA substrate within the sequence of the DNAzyme and simultaneously flanking it with the fluorophore and quencher, *cis*-acting RFDs were constructed.[32] These constructs could subsequently enter *in vitro* selection in order to isolate DNAzymes capable of recognizing and binding a specific molecular target with high affinity and specificity. Coupling RFD structure with *in vitro* selection ultimately yields a DNAzyme construct capable of molecular recognition-induced self-cleavage followed by emission of a fluorescent signal, indicating the presence of a specific molecular target in the solution.[32] Such constructs are the basis for this thesis.

The application of *in vitro* selection soon crossed over into the world of medical diagnostics and the advantages of functional nucleic acids were harnessed to perform targeted searches of biomarkers, resulting in numerous biomarker discoveries made across multiple clinical domains,

illustrated in Table 1.[33–42] Of note, since the start of the work described in this thesis, another group has attempted *in vitro* selection in human serum and successfully identified an aptamer for a biomarker associated with lung cancer.[43] To date, no studies exploring *in vitro* selection using DNAzymes appear to have been attempted for detection of biological targets.

Biomarker	Pathology	Method	Reference
Cyclophilin B (CypB)	Pancreatic cancer	Secretome SELEX	[44]
Heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1)	Breast cancer	Tissue SELEX	[45]
Tenascin-C	Glioblastoma	Cell-SELEX	[46]
Tyrosine kinase 7 (PTK7)	Leukaemia	Cell-SELEX	[47]
Stress-induced phosphoprotein 1 (STIP1)	Ovarian cancer	Cell-SELEX	[48]
Alkaline phosphatase placental-like 2 (ALPPL-2)	Pancreatic cancer	Cell-SELEX	[49]
Mortalin	Pancreatic cancer	Cell-SELEX	[50]
Vimentin	Cancer Metastasis	Cell-SELEX	[51]
High density lipoprotein binding protein (HDLBP)	Lung cancer	Cell-SELEX	[52]
Chemokine ligand 9 (CXCL9)	Transplant rejection	Split–combine click-SELEX	[53]
Tetranectin (CLEC3B)	Lung cancer	Serum SELEX	[43]

Table 1. Literature examples of disease biomarkers discovered using SELEX.

Traditional biomarker identification methods involve broad searches of biological samples with expensive equipment and techniques, including as gas chromatography-mass spectrometry, protein arrays, or extensive Next-Generation Sequencing.[54] These resource-intensive and time-consuming methods lack the finesse of *in vitro* selection and its ability to hone in on a target and perform a directed search for a biomarker. *in vitro* selection essentially functions as a molecular scanner, identifying molecular differences between biological samples, allowing researchers to select a DNAzyme probe which reacts only in the specific sample of interest. Once a DNAzyme probe is isolated, the biomarker causing the activation of the DNAzyme can be identified, simultaneously yielding (i) a sensitive and specific DNAzyme probe, and (ii) a potentially novel

biomarker.[55] The DNAzyme probe, by virtue of its high sensitivity and specificity to its target, along with prolonged intrinsic stability and low manufacturing cost, makes it an optimal choice for inclusion in biosensing platforms or assays for immediate clinical application.[56]

Furthermore, our research group has made significant contributions in the area of DNAzyme selection.[57–61] In recent years, our group has successfully employed this method to derive bacteria-specific DNAzymes from a synthetic DNA pool using *in vitro* selection to engineer high recognition specificity for a target of interest.[32] This strategy has been successfully applied to identify DNAzymes specific for *E. coli*, *C. difficile,* and *H. pylori* bacteria strains by virtue of a high affinity/high fidelity interaction between the selected DNAzyme and a strain-specific molecule of interest.[62–64] A similar technique will be adopted for DNAzyme selection in this thesis, since we believe this is a superior method of diagnostic tool development and biomarker detection.

1.4 A brief look at bioinformatics and *in vitro* selection

Steady advancements have been made with regards to DNA sequencing tools and methods.[65] Consequently, these advances have facilitated a partnership of *in vitro* selection and Next-Generation Sequencing, such that the progress of *in vitro* selection can be monitored through regular sequencing of DNA libraries recovered from distinct rounds of selection.[66] Yet, a vast amount of information provided by this elegant partnership remains unharvested.[66] Fortunately, with advancements in the field of bioinformatics, we can now access data previously unknown or deemed arcane.[67] This section will provide a brief summary of the evolving field of bioinformatics as it assists *in vitro* selection in enhancing data extraction from Next-Generation Sequencing data.

Next-Generation Sequencing (NGS) is a powerful tool capable of generating colossal datasets which require advanced bioinformatic analysis to organize and understand.[68] Consequently, upon sending a recovered DNA library for NGS and receipt of sequencing data, the first step of bioinformatic analysis consists of organization of the sequencing data generated.[69,70] Consequently, tools such as ClustalW and MUSCLE have been developed to facilitate sequence alignment while simultaneously scanning the quality of the recovered sequence reads, and eliminate low-quality reads.[71] Furthermore, these tools can be employed to quantify the copy numbers of each unique sequence in the dataset, yielding their frequency or abundance in the sequenced DNA library.[72] Equipped with this information, researchers can evaluate the diversity of the library and track its enrichment through numerical values or through enrichment plot visualization.[72] Additionally, both ClustalW and MUSCLE can be used to readily identify regions of conserved sequence identity, possibly indicative of structural or functional importance.[73] Such motifs can be further delineated using algorithms like MEME (Multiple Expectation maximizations for Motif Elicitation) and DREME (Discriminative Regular Expression Motif Elicitation).[74]

Following an assessment of the overall DNA library, bioinformatic algorithms can be employed to create clusters of sequences based on criteria of interest such as sequence homology.[74,75] In doing so, structural components yielded by conserved regions can be assessed against functionality and sequence frequency to determine promising sequence candidates selective for the target studied.[70,75] Based on the identity of a candidate sequence, bioinformatic tools such as RNAfold and mfold can predict the secondary structure of the sequence, providing researchers with possible insights into target-sequence interactions and susceptibility to degradation.[76]

Impressively, based on the sequencing data provided bioinformatic tools can further extrapolate the data and predict selection results in order to improve the efficiency of an *in vitro* selection protocol.[77] This process may take the form of bias mitigation by identifying whether the selection is hindered by unanticipated factors like primer binding or suboptimal amplification, and correcting for these biases.[78] Alternatively, this process may take the approach of applying bioinformatic models to simulate subsequent rounds of selection and predict the outcome of the protocol.[79] This tool offers researchers access to a further dimension of selection, namely *in silico* selection, aimed at mitigating limitations posed by lack of adequate time and resources.[79] Overall, the impact of bioinformatic tool development cannot be understated, particularly as it applies to enhancing the efficiency of *in vitro* selection. Consequently, only by integrating advanced bioinformatic analysis into the experimental workflow can we harness the true potential of *in vitro* selection.

1.5 Thesis objectives

The field of diagnostic and therapeutic tool development has taken enormous strides over the past century, advancing clinical medicine into the more advanced era of preventative medicine. Two pathologies with considerable associated morbidity and mortality have been profiled in this chapter. Yet despite the severity of both pancreatic cancer and malignant hyperthermia, there are no screening programs in place for early diagnosis and widespread prevention of morbidity and mortality associated with each of these conditions. In the world of preventative medicine, this reality is unacceptable.

Fortunately, advances in functional nucleic acid technology as it relates to diagnostic probe and tool development along with biomarker identification, can be harnessed towards the development of detection probes for each of these pathologies. Given that serum biomarkers carry the least invasive, most cost-effective, smallest side effect profile, and smallest risk of adverse outcomes compared to current diagnostic options available for each pathology, it reasons that future screening and detection assays would favour the use of serum biomarkers. Consequently, by applying a DNAzyme-based *in vitro* selection protocol for the first time directly to human serum, we are poised to be on the precipice of unlocking a large new chapter in the field of rare disease screening, in addition to establishing the feasibility of performing *in vitro* selection directly in whole human serum, and identifying potentially novel disease-specific biomarkers. Furthermore, recent advances in bioinformatic analysis tools will be explored to illustrate their ability to augment the efficacy of a high-risk selection protocol. Consequently, the objectives of this thesis aim to: (i) establish the feasibility of performing *in vitro* selection using DNAzymes directly in human serum, (ii) identify DNAzymes capable of detecting a pathology of interest, and (iii) develop bioinformatic strategies for evaluating the progress of an *in vitro* selection experiment and guide the selection of candidate sequences warranting further testing.

Two independent protocols have been designed and executed in Chapters 2 and 3, for the identification of DNAzyme sequences capable of detecting pancreatic cancer and malignant hyperthermia susceptibility, respectively. Control groups consisting of age and sex-matched patients with disease mimics were included to increase the specificity of the DNAzymes. Sensitivity and specificity analyses were conducted to assess the ability of the candidate sequences isolated to reliably detect the intended pathology. Finally, a size exclusion experiment was performed to identify possible target biomarker size parameters. Chapter 4 explores advances in

bioinformatic analysis and their application to *in vitro* selection sequencing results for evaluation of selection progress and efficiency, along with identification of promising DNAzyme sequences. Using these techniques, we demonstrate the emergence of a new candidate sequence capable of acting as a general cancer detection probe. Chapter 5 aims to provide a critical assessment of the application of *in vitro* selection in human serum, by exploring challenges encountered and offering conclusive explanations for discrepancies noted. Lastly, strategies to address the challenges discussed are provided, along with suggestions for fine-tuning the method for future work in this field. The final chapter of this thesis summarizes the achievements attained in this work.

1.6 References

1. Pereira MA, Chio IIC. Metastasis in pancreatic ductal adenocarcinoma: Current standing and methodologies. *Genes (Basel)*. 2020;11(1). doi:10.3390/genes11010006
2. Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. *The Lancet*. 2016;388(10039):73-85. doi:10.1016/S0140-6736(16)00141-0
3. Canadian Cancer Statistics Advisory in collaboration with the Canadian Cancer Society SC and the PHA of C. *Canadian Cancer Statistics: A 2022 Special Report on Cancer Prevalence*.; 2022.
4. Duffy MJ, Sturgeon C, Lamerz R, et al. Tumor markers in pancreatic cancer: A European Group on Tumor Markers (EGTM) status report. *Annals of Oncology*. 2009;21(3):441-447. doi:10.1093/annonc/mdp332
5. Progress M, Hidalgo M. Pancreatic Cancer. 2013;980:1605-1617. doi:10.1007/978-1-62703-287-2
6. Center MSKC. Pancreatic Cancer Diagnosis. Published 2018. Accessed August 20, 2011. https://www.mskcc.org/cancer-care/types/pancreatic/diagnosis
7. Brody JR, Witkiewicz AK, Yeo CJ. The Past, Present, and Future of Biomarkers: A Need for Molecular Beacons for the Clinical Management of Pancreatic Cancer. *Adv Surg*. 2011;45(1):301-321. doi:10.1016/j.yasu.2011.04.002

8. Koopmann J, Buckhaults P, Brown DA, et al. Serum Macrophage Inhibitory Cytokine 1 as a Marker of Pancreatic and Other Periampullary Cancers. *Clinical Cancer Research*. 2004;10(7):2386-2392. doi:10.1158/1078-0432.CCR-03-0165
9. Jones PM, Allen BN, Cherry RA, et al. Association between known or strongly suspected malignant hyperthermia susceptibility and postoperative outcomes: an observational population-based study. *Canadian Journal of Anesthesia*. Published online 2019. doi:10.1007/s12630-018-1250-8
10. Rosenberg H, Davis M, James D, Pollock N, Stowell K. Malignant hyperthermia. *Orphanet J Rare Dis*. 2007;2(1). doi:10.1186/1750-1172-2-21
11. Biesecker LG, Dirksen RT, Girard T, et al. Genomic Screening for Malignant Hyperthermia Susceptibility. *Anesthesiology*. 2020;133(6):1277-1282. doi:10.1097/ALN.0000000000003547
12. Rosenberg H, Pollock N, Schiemann A, Bulger T, Stowell K. Malignant hyperthermia: a review. *Orphanet J Rare Dis*. 2015;10(1):93. doi:10.1186/s13023-015-0310-1
13. Riazi S, Kraeva N, Hopkins PM. Updated guide for the management of malignant hyperthermia. *Canadian Journal of Anesthesia*. Published online 2018. doi:10.1007/s12630-018-1108-0
14. Riazi S, Larach MG, Hu C, Wijeysundera D, Massey C, Kraeva N. Malignant hyperthermia in Canada: Characteristics of index anesthetics in 129 malignant hyperthermia susceptible probands. *Anesth Analg*. Published online 2014. doi:10.1213/ANE.0b013e3182937d8b
15. Rosenberg H, Pollock N, Schiemann A, Bulger T, Stowell K. Malignant hyperthermia: a review. *Orphanet J Rare Dis*. 2015;10(1):93. doi:10.1186/s13023-015-0310-1
16. Bojko B, Vasiljevic T, Boyaci E, et al. Untargeted metabolomics profiling of skeletal muscle samples from malignant hyperthermia susceptible patients. *Canadian Journal of Anesthesia*. 2021;68(6):761-772. doi:10.1007/s12630-020-01895-y
17. Bandschapp O, Iaizzo PA, Girard T. Malignant hyperthermia – Update of diagnostics. *Trends in Anaesthesia and Critical Care*. 2012;2(5):218-223. doi:10.1016/j.tacc.2012.08.001

18. Litman RS, Griggs SM, Dowling JJ, Riazi S. Malignant Hyperthermia Susceptibility and Related Diseases. *Anesthesiology*. 2018;128(1):159-167. doi:10.1097/ALN.0000000000001877
19. Riazi S, Kraeva N, Hopkins PM. Malignant Hyperthermia in the Post-Genomics Era. *Anesthesiology*. 2018;128(1):168-180. doi:10.1097/ALN.0000000000001878
20. Riazi S, Kraeva N, Hopkins PM. Malignant Hyperthermia in the Post-Genomics Era. *Anesthesiology*. 2018;128(1):168-180. doi:10.1097/ALN.0000000000001878
21. McConnell EM, Cozma I, Mou Q, Brennan JD, Lu Y, Li Y. Biosensing with DNAzymes. *Chem Soc Rev*. 2021;50(16):8954-8994. doi:10.1039/D1CS00240F
22. Chang T, He S, Amini R, Li Y. Functional Nucleic Acids Under Unusual Conditions. *ChemBioChem*. 2021;22(14):2368-2383. doi:10.1002/cbic.202100087
23. Mcconnell EM, Cozma I, Morrison D, Li Y. Biosensors made of synthetic functional nucleic acids toward better human health. *Anal Chem*. 2020;92(1):327-344. doi:10.1021/acs.analchem.9b04868
24. Andrea L. Edwards, Robert T. Batey. Riboswitches: A Common RNA Regulatory Element. *Nature Education*. 2010;3(9):9-9.
25. Le Vay K, Mutschler H. The difficult case of an RNA-only origin of life. *Emerg Top Life Sci*. 2019;3(5):469-475. doi:10.1042/ETLS20190024
26. Gold L. SELEX: How It Happened and Where It will Go. *J Mol Evol*. 2015;81(5-6):140-143. doi:10.1007/s00239-015-9705-9
27. Wilson DS, Szostak JW. In Vitro Selection of Functional Nucleic Acids. *Annu Rev Biochem*. 1999;68(1):611-647. doi:10.1146/annurev.biochem.68.1.611
28. Zhang Y, Lai B, Juhas M. Recent Advances in Aptamer Discovery and Applications. *Molecules*. 2019;24(5):941. doi:10.3390/molecules24050941
29. Breaker RR, Joyce GF. A DNA enzyme that cleaves RNA. *Chem Biol*. 1994;1(4):223-229. doi:10.1016/1074-5521(94)90014-0
30. Tram K, Kanda P, Li Y. Lighting Up RNA-Cleaving DNAzymes for Biosensing. *J Nucleic Acids*. 2012;2012:1-8. doi:10.1155/2012/958683
31. Liu M, Chang D, Li Y. Discovery and Biosensing Applications of Diverse RNA-Cleaving DNAzymes. *Acc Chem Res*. 2017;50(9):2273-2283. doi:10.1021/acs.accounts.7b00262

32. Ali MM, Aguirre SD, Lazim H, Li Y. Fluorogenic DNAzyme Probes as Bacterial Indicators. *Angewandte Chemie International Edition*. 2011;50(16):3751-3754. doi:10.1002/anie.201100477
33. Bing T, Zhang N, Shangguan D. Cell-SELEX, an Effective Way to the Discovery of Biomarkers and Unexpected Molecular Events. *Adv Biosyst*. Published online 2019. doi:10.1002/adbi.201900193
34. Duan N, Wu S, Chen X, et al. Selection and Characterization of Aptamers against Salmonella typhimurium Using Whole-Bacterium Systemic Evolution of Ligands by Exponential Enrichment (SELEX). *J Agric Food Chem*. 2013;61(13):3229-3234. doi:10.1021/jf400767d
35. Taneja V, Goel M, Shankar U, et al. An Aptamer Linked Immobilized Sorbent Assay (ALISA) to Detect Circulatory IFN-α, an Inflammatory Protein among Tuberculosis Patients. *ACS Comb Sci*. Published online 2020. doi:10.1021/acscombsci.0c00108
36. Rotoli D, Santana-Viera L, Ibba ML, Esposito CL, Catuogno S. Advances in Oligonucleotide Aptamers for NSCLC Targeting. *Int J Mol Sci*. 2020;21(17):6075. doi:10.3390/ijms21176075
37. Nuzzo S, Brancato V, Affinito A, Salvatore M, Cavaliere C, Condorelli G. The role of RNA and DNA aptamers in glioblastoma diagnosis and therapy: A systematic review of the literature. *Cancers (Basel)*. Published online 2020. doi:10.3390/cancers12082173
38. Cheung YW, Röthlisberger P, Mechaly AE, et al. Evolution of abiotic cubane chemistries in a nucleic acid aptamer allows selective recognition of a malaria biomarker. *Proceedings of the National Academy of Sciences*. 2020;117(29):16790-16798. doi:10.1073/pnas.2003267117
39. Bing T, Wang J, Shen L, Liu X, Shangguan D. Prion Protein Targeted by a Prostate Cancer Cell Binding Aptamer, a Potential Tumor Marker? *ACS Appl Bio Mater*. 2020;3(5):2658-2665. doi:10.1021/acsabm.0c00024
40. SUN M, YANG G, ZHAO Y, QU F. Screening of Aptamer for Breast Cancer Biomarker Calreticulin and Its Application to Detection of Serum and Recognition of Breast Cancer Cell. *Chinese Journal of Analytical Chemistry*. 2020;48(5):642-649. doi:10.1016/S1872-2040(20)60020-2

41. Yoon S, Li H, Quintanar L, Armstrong B, Rossi JJ. Uncovering Differently Expressed Markers and Heterogeneity on Human Pancreatic Cancer. *Transl Oncol*. Published online 2020. doi:10.1016/j.tranon.2020.100749
42. Vargas-Montes M, Cardona N, Moncada DM, Molina DA, Zhang Y, Gómez-Marín JE. Enzyme-Linked Aptamer Assay (ELAA) for Detection of Toxoplasma ROP18 Protein in Human Serum. *Front Cell Infect Microbiol*. Published online 2019. doi:10.3389/fcimb.2019.00386
43. Zhao Y, He L, Huang B, et al. Identification of a novel DNA aptamer that selectively targets lung cancer serum. *RSC Adv*. 2021;11(53):33759-33769. doi:10.1039/D1RA06233F
44. Ray P, Rialon-Guevara KL, Veras E, Sullenger BA, White RR. Comparing human pancreatic cell secretomes by in vitro aptamer selection identifies cyclophilin B as a candidate pancreatic cancer biomarker. *Journal of Clinical Investigation*. 2012;122(5):1734-1741. doi:10.1172/JCI62385
45. Li S, Xu H, Ding H, et al. Identification of an aptamer targeting hnRNP A1 by tissue slide-based SELEX. *J Pathol*. 2009;218(3):327-336. doi:10.1002/path.2543
46. Daniels DA, Chen H, Hicke BJ, Swiderek KM, Gold L. A tenascin-C aptamer identified by tumor cell SELEX: systematic evolution of ligands by exponential enrichment. *Proc Natl Acad Sci U S A*. 2003;100(26):15416-15421. doi:10.1073/pnas.2136683100
47. Shangguan D, Cao Z, Meng L, et al. Cell-Specific Aptamer Probes for Membrane Protein Elucidation in Cancer Cells. *J Proteome Res*. 2008;7(5):2133-2139. doi:10.1021/pr700894d
48. Van Simaeys D, Turek D, Champanhac C, et al. Identification of Cell Membrane Protein Stress-Induced Phosphoprotein 1 as a Potential Ovarian Cancer Biomarker Using Aptamers Selected by Cell Systematic Evolution of Ligands by Exponential Enrichment. *Anal Chem*. 2014;86(9):4521-4527. doi:10.1021/ac500466x
49. Dua P, Kang HS, Hong SM, Tsao MS, Kim S, Lee D ki. Alkaline Phosphatase ALPPL-2 Is a Novel Pancreatic Carcinoma-Associated Protein. *Cancer Res*. 2013;73(6):1934-1945. doi:10.1158/0008-5472.CAN-12-3682

50. Li Y, TAM WW, Yu Y, et al. The application of Aptamer in biomarker discovery. *Biomark Res*. 2023;11(1):70. doi:10.1186/s40364-023-00510-8
51. Yoon S, Armstrong B, Habib N, Rossi JJ. Blind SELEX Approach Identifies RNA Aptamers That Regulate EMT and Inhibit Metastasis. *Molecular Cancer Research*. 2017;15(7):811-820. doi:10.1158/1541-7786.MCR-16-0462
52. Zhou W, Zhao L, Yuan H, et al. A new small cell lung cancer biomarker identified by Cell-SELEX generated aptamers. *Exp Cell Res*. 2019;382(2):111478. doi:10.1016/j.yexcr.2019.06.023
53. Siegl J, Nikolin C, Phung NL, Thoms S, Blume C, Mayer G. Split–Combine Click-SELEX Reveals Ligands Recognizing the Transplant Rejection Biomarker CXCL9. *ACS Chem Biol*. 2022;17(1):129-137. doi:10.1021/acschembio.1c00789
54. Tothill I, Altintas Z. Molecular biosensors: promising new tools for early detection of cancer. *Nanobiosensors in Disease Diagnosis*. 2015;(January):1. doi:10.2147/NDD.S56772
55. Cozma I, McConnell EM, Brennan JD, Li Y. DNAzymes as key components of biosensing systems for the detection of biological targets. *Biosens Bioelectron*. 2021;177:112972. doi:10.1016/j.bios.2021.112972
56. Ma L, Liu J. Catalytic Nucleic Acids: Biochemistry, Chemical Biology, Biosensors, and Nanotechnology. *iScience*. 2020;23(1):100815. doi:10.1016/j.isci.2019.100815
57. Mei SHJ, Liu Z, Brennan JD, Li Y. An efficient RNA-cleaving DNA enzyme that synchronizes catalysis with fluorescence signaling. *J Am Chem Soc*. 2003;125(2):412-420. doi:10.1021/ja0281232
58. Manochehry S, Gu J, McConnell EM, Salena BJ, Li Y. In Vitro Selection of New DNA Aptamers for Human Vascular Endothelial Growth Factor 165. *ChemBioChem*. 2020;21(14):2029-2036. doi:10.1002/cbic.202000024
59. Liu M, Yin Q, Brennan JD, Li Y. Selection and characterization of DNA aptamers for detection of glutamate dehydrogenase from Clostridium difficile. *Biochimie*. Published online 2018. doi:10.1016/j.biochi.2017.08.015
60. Zhang A, Chang D, Zhang Z, et al. In vitro selection of DNA aptamers that binds geniposide. *Molecules*. Published online 2017. doi:10.3390/molecules22030383

61. Zhang W, Feng Q, Chang D, Tram K, Li Y. In vitro selection of RNA-cleaving DNAzymes for bacterial detection. *Methods*. 2016;106:66-75. doi:10.1016/j.ymeth.2016.03.018
62. Ali MM, Aguirre SD, Lazim H, Li Y. Fluorogenic DNAzyme Probes as Bacterial Indicators. *Angewandte Chemie International Edition*. 2011;50(16):3751-3754. doi:10.1002/anie.201100477
63. Aguirre S, Ali M, Salena B, Li Y. A Sensitive DNA Enzyme-Based Fluorescent Assay for Bacterial Detection. *Biomolecules*. 2013;3(4):563-577. doi:10.3390/biom3030563
64. Liu M, Chang D, Li Y. Discovery and Biosensing Applications of Diverse RNA-Cleaving DNAzymes. *Acc Chem Res*. 2017;50(9):2273-2283. doi:10.1021/acs.accounts.7b00262
65. Bayat A. Science, medicine, and the future: Bioinformatics. *BMJ*. 2002;324(7344):1018-1022. doi:10.1136/bmj.324.7344.1018
66. Schütze T, Wilhelm B, Greiner N, et al. Probing the SELEX Process with Next-Generation Sequencing. *PLoS One*. 2011;6(12):e29604. doi:10.1371/journal.pone.0029604
67. Blank M. Next-Generation Analysis of Deep Sequencing Data: Bringing Light into the Black Box of SELEX Experiments. In: ; 2016:85-95. doi:10.1007/978-1-4939-3197-2_7
68. Schmidt B, Hildebrandt A. Next-generation sequencing: big data meets high performance computing. *Drug Discov Today*. 2017;22(4):712-717. doi:10.1016/j.drudis.2017.01.014
69. Madeira F, Park Y mi, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019;47(W1):W636-W641. doi:10.1093/nar/gkz268
70. Weiss Z, DasGupta S. REVERSE: a user-friendly web server for analyzing next-generation sequencing data from *in vitro* selection/evolution experiments. *Nucleic Acids Res*. 2022;50(W1):W639-W650. doi:10.1093/nar/gkac508
71. Fosso B, Santamaria M, Marzano M, et al. BioMaS: a modular pipeline for Bioinformatic analysis of Metagenomic AmpliconS. *BMC Bioinformatics*. 2015;16(1):203. doi:10.1186/s12859-015-0595-z

72. Hoinka J, Backofen R, Przytycka TM. AptaSUITE: A Full-Featured Bioinformatics Framework for the Comprehensive Analysis of Aptamers from HT-SELEX Experiments. *Mol Ther Nucleic Acids*. 2018;11:515-517. doi:10.1016/j.omtn.2018.04.006
73. Gao S, Zheng X, Jiao B, Wang L. Post-SELEX optimization of aptamers. *Anal Bioanal Chem*. 2016;408(17):4567-4573. doi:10.1007/s00216-016-9556-2
74. Pei S, Slinger BL, Meyer MM. Recognizing RNA structural motifs in HT-SELEX data for ribosomal protein S15. *BMC Bioinformatics*. 2017;18(1):298. doi:10.1186/s12859-017-1704-y
75. Kato S, Ono T, Minagawa H, et al. FSBC: fast string-based clustering for HT-SELEX data. *BMC Bioinformatics*. 2020;21(1):263. doi:10.1186/s12859-020-03607-1
76. Rodriguez LFR, Cortes-Mancera FM. Evaluation of nine bioinformatic platforms using a viral model for RNA secondary structure prediction. In: *Symposium of Signals, Images and Artificial Vision - 2013: STSIVA - 2013*. IEEE; 2013:1-5. doi:10.1109/STSIVA.2013.6644925
77. Beier R, Boschke E, Labudde D. New Strategies for Evaluation and Analysis of SELEX Experiments. *Biomed Res Int*. 2014;2014:1-12. doi:10.1155/2014/849743
78. Kohlberger M, Gadermaier G. SELEX: Critical factors and optimization strategies for successful aptamer selection. *Biotechnol Appl Biochem*. 2022;69(5):1771-1792. doi:10.1002/bab.2244
79. Ahmad NA, Mohamed Zulkifli R, Hussin H, Nadri MH. In silico approach for Post-SELEX DNA aptamers: A mini-review. *J Mol Graph Model*. 2021;105:107872. doi:10.1016/j.jmgm.2021.107872

Chapter 2. Identification of RNA-cleaving Fluorogenic DNAzymes for Detection of Pancreatic Cancer in Human Serum

2.1 Introduction

Functional nucleic acid technology has been dutifully employed and perfected in our research group in order to create diagnostic solutions for point-of-care testing.[1–5] The work in this chapter aims to add to the repertoire of detection probes isolated by our group, by extending the search to biological targets of disease in biologic samples. As such, the focus of this chapter is the derivation and application of RNA-cleaving fluorogenic DNAzymes (RFDs) through *in vitro* selection, towards the detection of pancreatic cancer in human serum.

Despite the versatility of *in vitro* selection and the wealth of metabolic substrates available in human serum, at the start of this work no other groups had attempted *in vitro* selection directly in human serum towards the identification of biologic targets. Consequently, we endeavoured to determine whether such an exploration would be feasible given the immense potential for diagnostic tool development and biomarker identification. Furthermore, given the dire need for an accessible, inexpensive diagnostic and screening tool for early detection of pancreatic cancer, we opted to focus our efforts on the identification of DNAzymes capable of specifically recognizing pancreatic cancer in human serum.

Through repetitive exposure to a target of interest, DNAzymes can be trained to recognize the target and simultaneously perform a catalytic activity such as cleavage of a substrate to indicate binding/interaction with the target.[6] In order to make DNAzymes function optimally for *in vitro* selection, they must be specifically engineered to optimally harness their catalytic activity. Specific consideration is therefore given to the design of the DNAzyme sequence, bearing in mind

the DNAzyme must have a target-interacting section within its sequence, a catalytic section, and a signal generation section as a measure of its activity.[7]

RNA-cleaving DNAzymes are able to cleave a substrate at a specific RNA dinucleotide junction in response to a stimulus, such as binding/interaction with a target of interest.[8] By embedding the ribonucleotide junction within the sequence of the DNAzyme, we have constructed a self-cleaving DNAzyme. Upon interaction with a molecular target, the DNAzyme becomes activated and proceeds to self-cleave specifically and predictably at the same site, for ease of measurement and detection of cleavage. Cleaved DNAzymes will be identified by comparing their size pre- and post-serum incubation via gel electrophoresis, simultaneously purifying the cleaved sequences, and concluding one round of selection.[8]

In order to quantify the amount of self-cleavage observed within a large pool of DNAzymes, each RNA-cleaving DNAzyme is fashioned with a signal-generating element. This element will be referred to as FQ30. FQ30 is a thirty-oligonucleotide long sequence which can be ligated to the RNA-cleaving DNAzyme. It contains a fluorophore and quencher moiety separated by 1-2 oligonucleotides. When in close proximity to each other, the quencher moiety absorbs the fluorescence generated by the fluorophore moiety, and little detectable fluorescence is emitted. Separation of the fluorophore and quencher therefore allows the emission of large, readily detectable levels of fluorescence. Therefore, by embedding the RNA junction (DNAzyme cleavage site) within the FQ30 sequence, between the fluorophore and quencher, and ligating this FQ30 sequence to the DNAzyme, we have effectively engineered a self-cleaving DNAzyme capable of simultaneously recognizing a target of interest, cleaving itself upon doing so, and generating a

detectable, quantifiable fluorescence signal to indicate whether cleavage has occurred. This concept is illustrated in Figure 1.

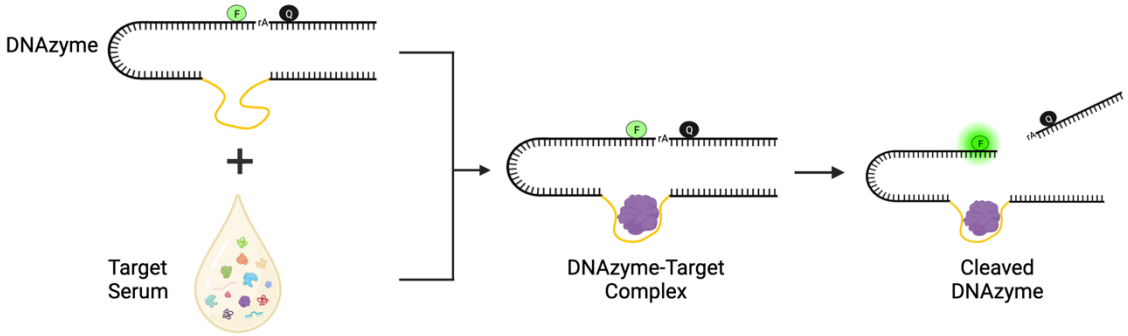


Figure 1. Conceptual design of an RNA-cleaving fluorogenic DNAzyme interacting with a target of interest in human serum, and emitting a detectable fluorescent signal upon cleavage.

Careful consideration must also be given to the internal structure of the DNAzyme – both for maximizing the diversification of sequences, and ease of sequence amplification. Given the substantial complexity of human serum, and the large number of unknown potential targets within, the number of unique DNAzyme sequences must be maximized. Doing so increases the likelihood of DNAzyme-target pairings to emerge. The variability of the DNAzyme sequences can be achieved through the incorporation of a lengthy variable domain, henceforth named the random domain. The random domain enables the development of a DNA library, with quadrillions of permutations of nucleotides in the random domain. Our research group has previously developed a model random sequence DNA library consisting of 10^{16} unique sequences of DNA, by virtue of a 40-nucleotide random domain. This model was the basis for the DNA library employed in this study.

As previously mentioned, sequences that behave in accordance with the *in vitro* selection pressures placed on the DNA library population are carried forward into the experiment via amplification in each round, achieved with real-time PCR. In order to facilitate amplification of favourable sequences during each round of selection, the random domain is flanked by 2 constant regions, which serve as forward and reverse primers for PCR. Consequently, sequences that adhere to the selection pressures most readily are amplified early in the process and can be distinguished at the end of the experiment by their significant abundance compared to the other remaining sequences in the enriched library. The internal structure of the DNA library, FQ30 segment and architecture of an RFD is illustrated in Figure 2.

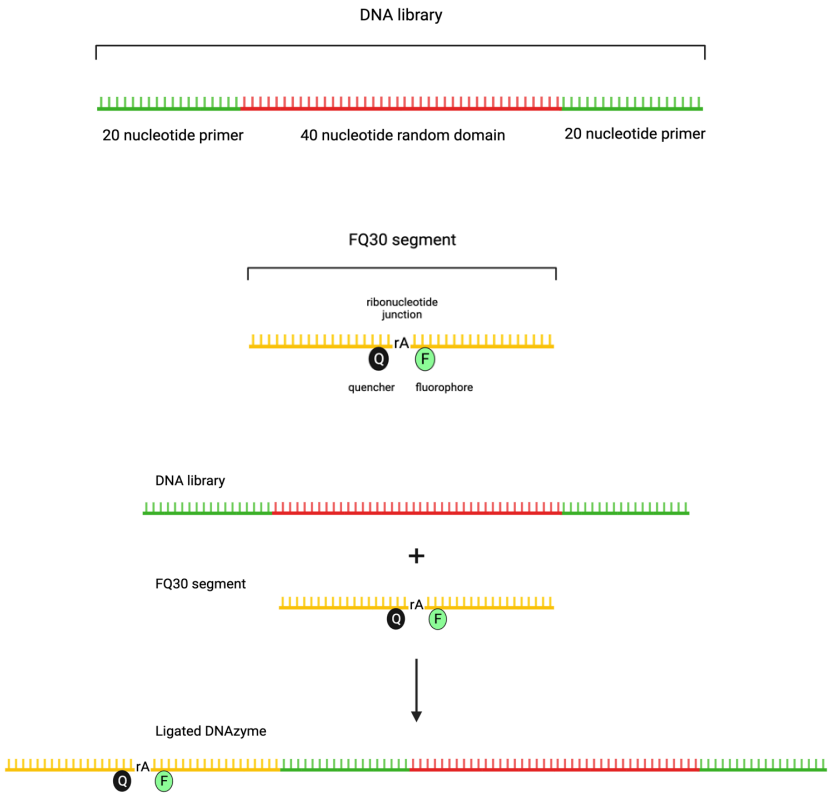


Figure 2. Conceptual design of a DNA library and FQ30 signal generating segment. Construction of RNA-cleaving fluorogenic DNAzyme, following ligation of the DNA library with the FQ30 segment.

2.1.1 Study design

This study was designed with two phases in mind. The first phase is defined as the preliminary phase, whereby the process of *in vitro* selection is applied to pooled serum samples from patients with pancreatic cancer, in the hopes of identifying a DNAzyme sequence sensitive and specific for pancreatic cancer detection. To achieve specificity from other cancer types, the control serum was designed to be exclusively formed by pooling serum from patients diagnosed with the four most common cancers by incidence – breast, colorectal, lung, and prostate.[9] By introducing potential overlap of biomarkers/serum components early in the selection, we avoid having to later account for low specificity. By tasking the DNA library early on with identifying biomarkers unique for pancreatic cancer, we are programming it to maximize its specificity early on.

The second phase of this project is defined as the validation stage, whereby any candidate DNAzyme sequences are tested in new serum samples to obtain a benchmark sensitivity and specificity profile. The validation phase requires recruitment of patients from local area hospitals, and is designed to include pancreatic cancer patients, other cancer patients, and healthy patients. Due to the unfortunate incidence of the COVID-19 pandemic, recruitment of patients was severely impacted and only 10 new patient samples could be collected for the validation phase.

The first phase of this project begins with the acquisition of preoperative/pre-treatment plasma samples from the Ontario Tumour Bank (OTB) for preliminary testing. The samples consisted of 10 pancreatic cancer plasma samples, and 10 plasma samples of each of the following cancer types for use in counter selection: breast, colorectal, lung, and prostate. All plasma samples required processing to obtain serum (seroconversion by defibrination). Following processing, the samples

were aliquoted into 20 µl volumes to minimize the number of freeze-thaw cycles experienced, which may degrade any potential target(s). Upon seroconversion and preparation of the serum samples, the *in vitro* selection protocol can be initiated as depicted in Figure 3.

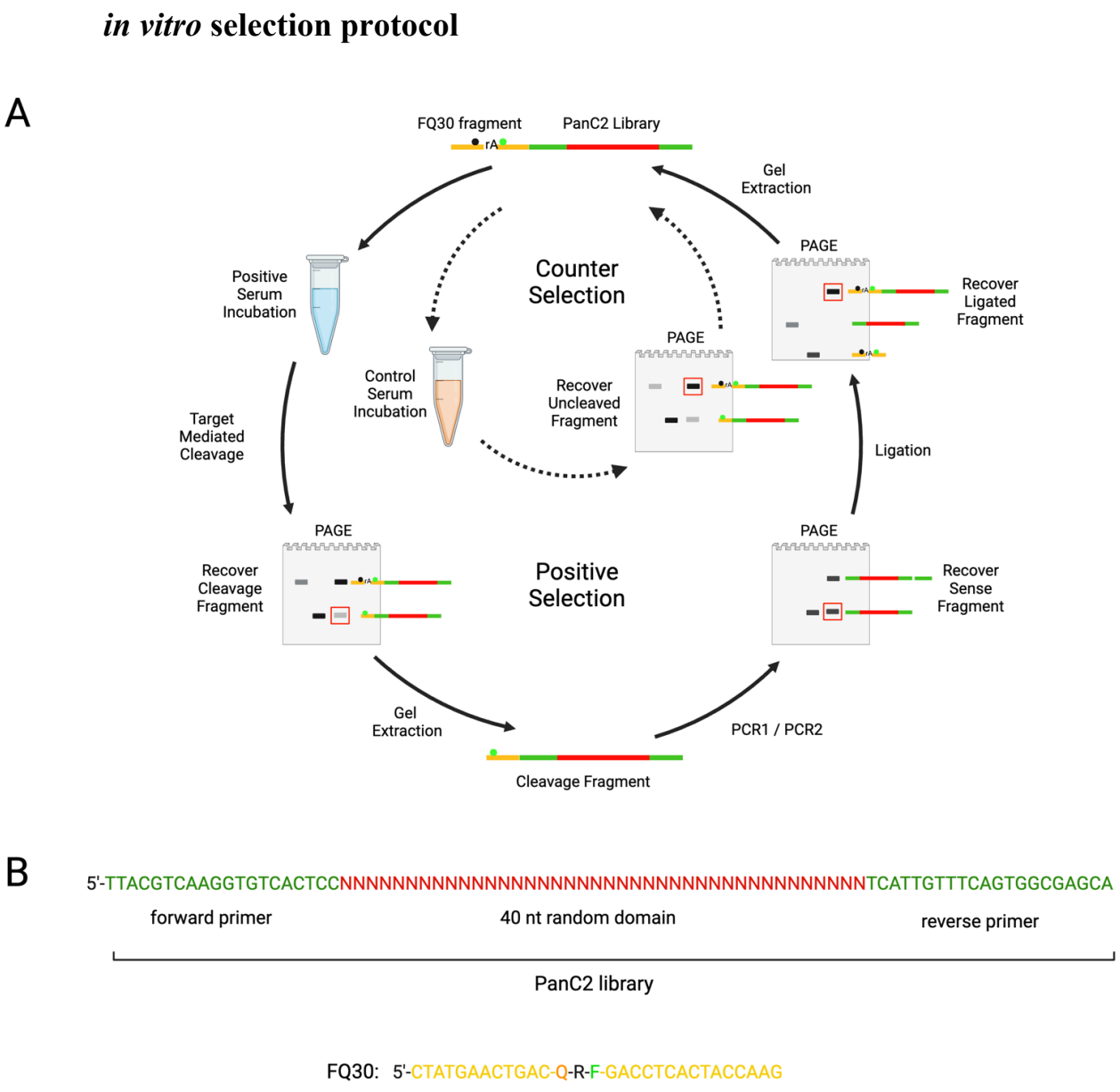


Figure 3. A. *in vitro* Selection/SELEX scheme using pancreatic cancer serum and the PanC2 DNA library. Counter and positive selection cycles depicted, performed in iterative series. B. Oligonucleotide sequence of DNA library and FQ30 substrate.

Beginning with the ligated PanC2 library, the aim of the selection protocol illustrated in Figure 1 is to guide the researcher through iterative rounds of Counter and Positive Selection. A sequential explanation of the figure is detailed below.

i. PanC2 library is ligated with the FQ30 substrate, which contains the fluorophore, ribonucleotide (site of DNAzyme cleavage), and quencher. The ligated product proceeds into selection.

ii. The ligated product from the ligation is incubated with Counter Selection Serum (serum of colon, lung, breast, and prostate cancer). Succeeding incubation (obeying counter selection conditions), the selection reaction is run on a Urea PAGE gel, in order to separate the cleaved and uncleaved sequences. The cleaved sequences are discarded (archived) and the uncleaved sequences are purified and carried over into positive selection.

iii. The purified uncleaved sequences from the Counter Selection are incubated with Positive Selection Serum (pancreatic cancer serum). Succeeding incubation (obeying positive selection conditions), the selection reaction is run on a Urea PAGE gel, in order to separate the cleaved and uncleaved sequences. The cleaved sequences are extracted from the gel and purified.

iv. Following purification of cleaved positive selection sequences, 2 rounds of PCR are performed. The first round serves to amplify and give a rough estimation of the amount of sequences obtained at the end of each round of selection. It also serves as a benchmark indicator for the required amplification cycles necessary until plateau. The second round of PCR serves to mass amplify the recovered sequences from each round of selection, tagging the antisense strands with an extension blocked primer marked by a poly-T tail (poly-thymine tail of 20 thymine nucleotides linked to the reverse primer by an 18 atom spacer).

v. Following PCR amplification, the PCR product is run on a Urea PAGE gel in order to separate the sense and antisense strands. Antisense strands will advance less on the gel due to their increased size from the addition of the extension blocked primer. The sense strands are extracted from the gel and purified.

vi. The purified sense strands which constitute the recovered enriched library once again undergo ligation to form DNAzymes, in preparation for the next round of Counter and Positive Selection.

Conditions of Counter and Positive Selection

Table 1 denotes the reaction conditions of the counter and positive selection reactions. The first round is proposed to start with 2000pmol FQ30 ligated library. In the interest of expediting the start of the selection protocol, we proceeded with 1300pmol ligated library. The library concentration in each selection reaction was kept constant at 500nM beginning at round 3 in order to avoid dimerization of DNAzyme sequences and formation of inappropriate complexes limiting functionality. The first 3 rounds consist of higher library input and concentration given the predicted low level of interaction between the DNAzyme library and the positive selection serum. Increasing the concentration increases the opportunity for the library to identify potential targets, while the larger input serves to facilitate recovery of cleaved sequences.

In addition, the selection conditions consist of a short incubation time interval for positive selection, and long incubation interval for counter selection. The short positive selection incubation time serves to push for the selection of high-affinity DNAzyme sequences, which translate to a highly selective and sensitive probe. This protocol also reflects the use of aggressive

counter selection measures. First, the frequency of counter selection must be noted. By including a round of counter selection before every round of positive selection, we can regularly filter out sequences that erroneously interact with targets in control serum, resulting in inappropriate cleavage, and effectively boost the specificity of the recovered library at the end of each round of selection. Second, the duration of incubation with counter control serum is 2x to 48x longer than with positive selection serum. Increasing the length of time the DNAzyme library is exposed to the control serum is another measure of increasing specificity, by giving ample amount of time for any non-specific or borderline specific sequences to identify a target in the control serum, leading to their removal from the library. Consequently, the combined incubation conditions for positive and counter selection ensure that the library recovered at the end of each round of selection is not only highly sensitive but also highly specific to the intended target and serum.

Throughout the selection, the library was sent for sequencing at benchmark intervals to determine the effectiveness of the selection pressures applied, track the progress of the selection, and choose DNAzyme candidates for further testing and evaluation. Consequently, candidate DNAzyme sequences underwent sensitivity and specificity analysis by performing cleavage time course reactions in individual patient samples (OTB and new) and observing whether any sequences reliably and preferentially cleave in pancreatic cancer serum, over serum of other cancer types, or healthy human serum.

Round	Round Type	[Target]	Target Sample	[Library] (nM)	Library Input (pmol)	Reaction Time (hours)
1	+	30%	Pancreatic Cancer Pool	5000	2000	1
2	-	1X	Selection Buffer	1000	200	2
2	+	30%	Pancreatic Cancer Pool	1000	100	1
3	-	1X	Selection Buffer	500	200	2
3	+	30%	Pancreatic Cancer Pool	500	100	1
4	-	30%	Counter Cancer Pool	500	200	2
4	+	30%	Pancreatic Cancer Pool	500	100	1
5	-	30%	Counter Cancer Pool	500	200	2
5	+	30%	Pancreatic Cancer Pool	500	100	1
6	-	30%	Counter Cancer Pool	500	200	2
6	+	30%	Pancreatic Cancer Pool	500	100	1
7	-	30%	Counter Cancer Pool	500	200	2
7	+	30%	Pancreatic Cancer Pool	500	100	1
8	-	30%	Counter Cancer Pool	500	200	2
8	+	30%	Pancreatic Cancer Pool	500	100	1
9	-	30%	Counter Cancer Pool	500	200	2
9	+	30%	Pancreatic Cancer Pool	500	100	1
10	-	30%	Counter Cancer Pool	500	200	2
10	+	30%	Pancreatic Cancer Pool	500	100	1
11	-	30%	Counter Cancer Pool	500	200	24
11	+	30%	Pancreatic Cancer Pool	500	100	1
12	-	30%	Counter Cancer Pool	500	200	24
12	+	30%	Pancreatic Cancer Pool	500	100	1
13	-	30%	Counter Cancer Pool	500	200	48
13	+	30%	Pancreatic Cancer Pool	500	100	1
14	-	30%	Counter Cancer Pool	500	200	48
14	+	30%	Pancreatic Cancer Pool	500	100	1

Table 1. Selection conditions for all 14 rounds of positive (+) and counter (-) selection using the PanC2 DNA library.

2.1.2 Sample procurement and processing

Ontario Tumor Bank

The conditions given to the OTB for sample selection included patients >18 years of age, pre-treatment and preoperative samples. Acquiring pre-treatment and preoperative samples serves to maximize the potential for biomarker discovery, since the samples will be unadulterated by chemotherapy or resection, which we predict would decrease the amount of circulating tumour markers. We additionally specified the need for early stage pancreatic cancer samples. Although this may compromise the tumour biomarker burden in the serum, and make it more challenging for our DNAzyme library to identify potential targets of interest, we insist on selecting a high-affinity DNAzyme capable of identifying early stage pancreatic cancer amenable to resection and curative ability. Successful identification of a highly sensitive and specific DNAzyme could be immediately used as a diagnostic and screening test. As can be observed in Table 2, one of the current limitations in basic science research is access to clinical samples. While efforts were made by OTB staff to select samples adhering to all of our conditions, not all samples conform to our criteria.

	Pancreatic Cancer	*Breast Cancer*	*Colorectal Cancer*	*Lung Cancer*	*Prostate Cancer*
Total Samples	10	10	10	10	10
Sex distribution	6M/4F	10F	7M/3F	5M/5F	10M
Age Range	35-79	50-89	40-94	25-84	45-74
Age Range of Samples (year of collection)	2010-2015	2009-2015	2009-2015	2006-2009	2010-2017
Cancer subtype	PDA (6) PNET (4)	Invasive Carcinoma	Invasive Carcinoma	Invasive Carcinoma	Intraepithelial neoplasia
Preoperative	10/10	9/10	9/10	10/10	10/10

Table 2. Specific clinical details of the samples obtained from the OTB. M = male. F = female. PDA = pancreatic ductal adenocarcinoma. PNET = pancreatic neuroendocrine tumour.

All OTB samples were received as plasma and underwent seroconversion through addition of 1M Calcium Chloride solution (in 2.5μmol doses) to initiate clotting. Calcium is a crucial cofactor in the blood clotting cascade[10], and was found to be a reliable method of seroconversion.[11] An added benefit of calcium stands in the avoidance of addition of exogenous targets (i.e. thrombin) which may provide a false target for DNAzymes to identify, confounding the selection. In cases where clotting was not stimulated an hour after the addition of Calcium Chloride, a second dose of was given. If no clotting was observed after a second dose of Calcium Chloride, 10 silica beads were added to stimulate clotting. All 50 OTB samples were successfully converted using 1-2 doses of Calcium Chloride and silica beads.

Local Patient Recruitment

Upon identification of potentially eligible patients, their charts were reviewed to ensure their eligibility. Specific details reviewed were the age and sex of the patient, the preoperative diagnosis, clinical tests and imaging. Based on these criteria, patients were preselected into our study and approached at their next available hospital appointment. The consent form designed for each study was reviewed with the patient and signed. Every patient was provided with a copy of the signed consent form, and the original was stored by the principal investigator.

Individual Patient Sample collection

The volume collected was ~5-10ml of whole blood. With the exception of one pancreatic cancer patient sample, all recruited patient samples were collected in red-capped blood collection tubes (Figure), which are free of chelating agents like EDTA or citrate, as well as other substances that prevent/delay blood clotting like heparin. Specific clinical details for the samples collected from local area patients are provided in Table 3.

	Pancreatic Cancer	*Healthy Participants*
Total Samples	5	5
Sex distribution	5M	2M/3F
Age Range	50-84	20-58
Age Range of Samples (year of collection)	2019-2020	2023
Cancer subtype	PDA (5)	N/A
Preoperative	5/5	N/A

Table 3. Specific clinical details of the samples obtained from patients recruited from local area hospitals.

Individual Patient Sample Processing

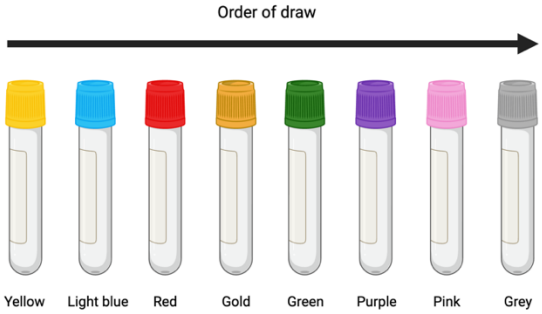
Serum Processing

Samples collected in red-capped vacutainer tubes were successfully converted to serum by allowing the blood to clot at room temperature, and centrifuged at 4°C. The recovered supernatant (serum) was transferred to cryovials, labelled and stored at -80°C.

Plasma Processing

One sample was collected in an EDTA-coated vacutainer, which required a two-step conversion to serum. First, the blood sample was centrifuged at 4°C to separate the blood into its three fractions – plasma, buffy coat, and red blood cells. The recovered supernatant (plasma) was transferred to cryovials, labelled and stored at -80°C. On demand seroconversion by defibrination was performed on fractions of the recovered plasma using 1M Calcium Chloride doses, akin to the defibrination of the OTB samples. The sample was successfully converted.

A



B

Colour Cap	Coating	Clinical Use
Yellow	Citrate	Blood culture
Light blue	3.2% sodium citrate	Coagulation studies
Red 1	No additive	Serum biochemistry
Red 2	Silica coating for clot activation	Serum biochemistry
Gold	Silica and gel for serum separation	Serum biochemistry
Green	Heparin	Plasma biochemistry
Purple	EDTA	Hematology
Pink	Potassium EDTA	Blood typing, cross-matching
Grey	Sodium Fluoride	Glucose estimation

C

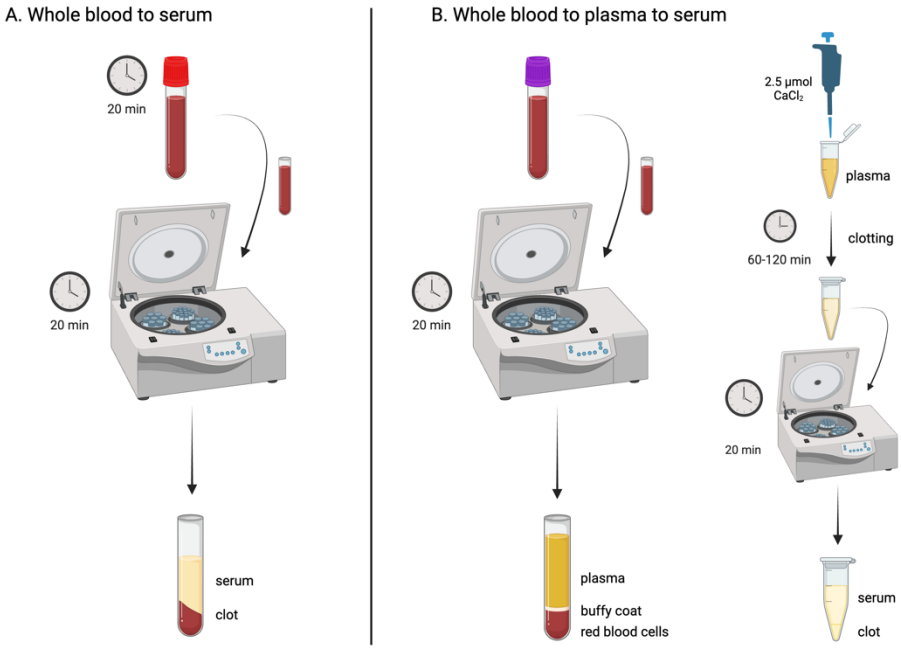


Figure 4. A. Blood collection vacutainers. **B.** Description of additive substance coating each type of tube and intended use. Serum sample preparation. **C.** Seroconversion from whole blood (left) and via plasma defibrination (right).

2.2 Results

2.2.1 *in vitro* selection

The PanC2 library underwent ligation with the FQ30 segment, yielding 1300pmol of ligated product. This product was carried into positive selection, followed by alternating rounds of counter and positive selection. Counter selection for rounds 2 and 3 was carried out with selection buffer to ensure any sequences that cleave in response to buffer/self-cleave without a target trigger are removed from the population. Beginning with round 4, counter selection was carried out with the counter selection pooled cancer serum. Each selection reaction was run on a PAGE gel, and the corresponding band was excised and purified. The cleavage bands and percent cleavage was calculated for every selection reaction, and the results of the gel cleavage analysis are outlined in Figure 5.

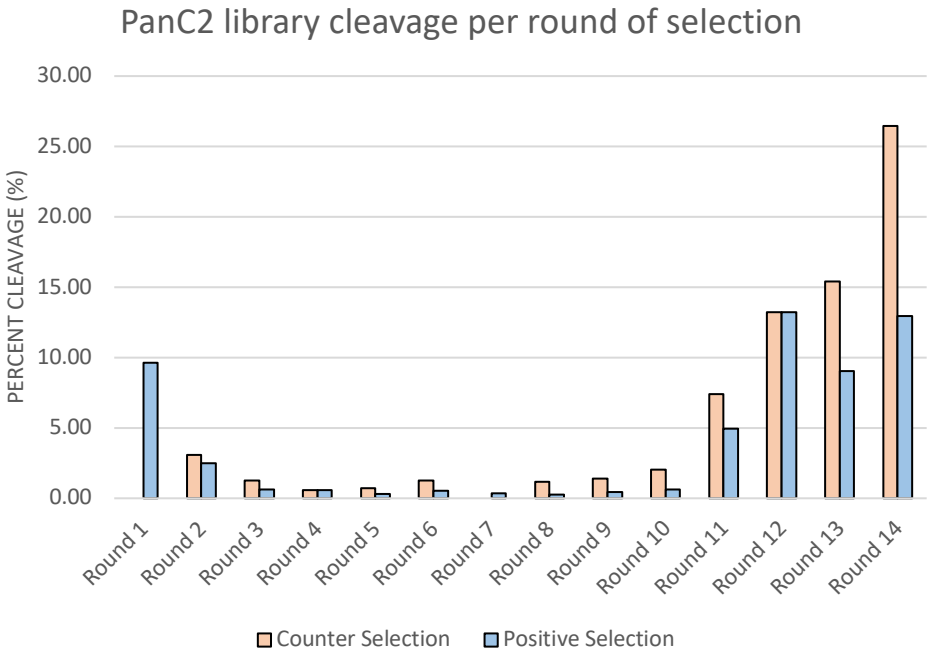


Figure 5. Graph summary of the percent cleavage of ligated library product for Rounds 1-14 of positive and counter selection.

Several notable trends have emerged from this analysis. First, across the first ten rounds of selection, both the positive and counter selection reactions average low DNAzyme library cleavage rates of approximately 1%. This finding is consistent with previous selections performed within our group, and reflects the high diversity of sequences in the population/library. Over the course of the selection, as the library becomes enriched with higher frequencies of sequences which respond to pancreatic cancer serum and not to the counter selection cancer pool, we expected to see an incremental increase in the DNAzyme population cleavage rates especially in the positive selection reactions. Consistent values at 1% did raise the question whether enrichment was indeed occurring. In order to address this concern, a cleavage time course experiment was conducted following the fifth and seventh round of selection. A cleavage time course reaction consists of an extended incubation of the DNAzyme library (or individual sequences) with a target of interest, with removal of aliquots of the reaction at specific time points to track the rate of cleavage and activity of the DNAzymes over time. The cleavage time course reactions were set up to compare the activity of the enriched library pool from the 5^{th} round of selection to the starting library (R0). The results of this comparative cleavage time course reaction are illustrated in Figure 6.

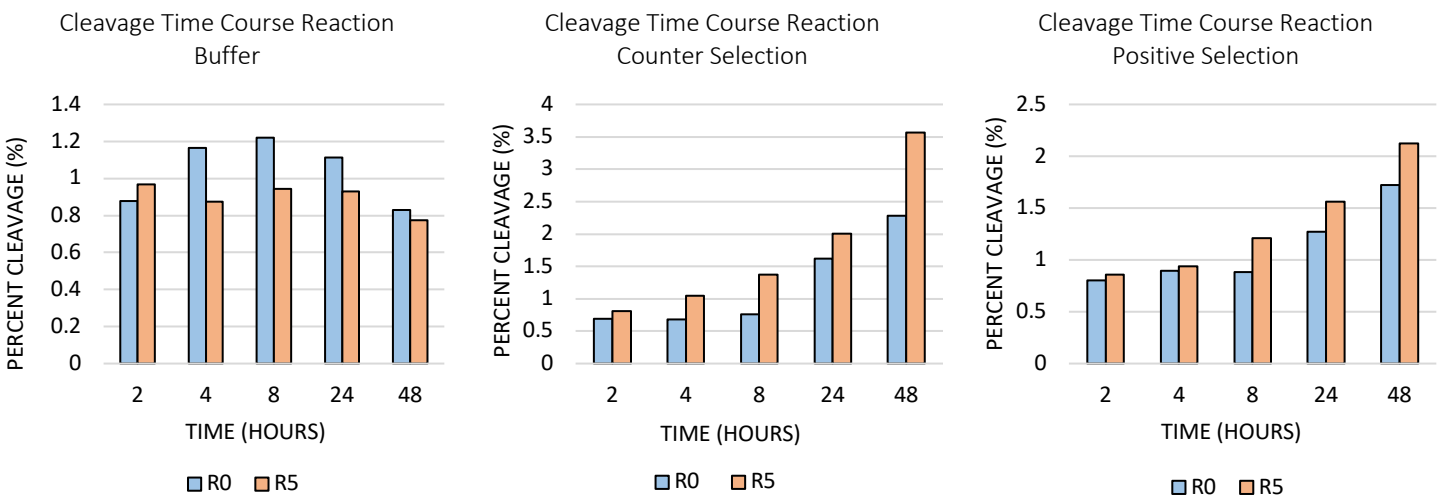


Figure 6. Comparison of Round 0 and the Round 5 recovered library performance in a cleavage time course reaction in 3 media.

The results of the cleavage time course show that the DNAzyme library is indeed enriching. Starting with the buffer reaction, it can be readily seen that the extent of cleavage decreases as we advance through rounds of selection. This indicates that we are effectively removing sequences which self-cleave without a target or react to buffer components from the library. This is an early indication that the selection is proceeding well. We do expect some background cleavage/degradation as part of the process of ligation and setting up the selection reactions; however, the fact that the cleavage percentages do not increase with time and are in fact remain consistently at or below 1% reassures us that we have created a DNAzyme library with high functional stability through 48 hours, and one that is not subject to acute active degradation.

The results of the cleavage time course in counter selection serum show that both libraries' cleavage levels increase over time, suggesting that longer incubation time likely correlates with more target interactions. Additionally, low levels of library cleavage can be observed at Round 0,

and comparatively higher levels of cleavage at Round 5. The increased cleavage levels at Round 5 indicate enrichment of the library, and a possible commonality in the targets recognized in counter selection serum and pancreatic cancer serum.

Lastly, the results of the cleavage time course in positive selection serum further indicate that enrichment of the DNAzyme library is occurring over the 5 selection rounds performed. Similar to the counter selection reactions, both show increasing cleavage levels over time, once again suggesting the degree of interaction of our library with potential targets is likely time dependent. Although modest, the Round 5 cleavage activity is higher than the activity of the starting library at every time point, demonstrating enrichment. Subsequently, additional rounds of selection are required to continue fine-tuning the library and increase its affinity for the positive selection serum. Ultimately, this cleavage time course reaction experiment establishes that enrichment of the DNAzyme library is occurring, an exciting early confirmation of the feasibility of performing *in vitro* selection directly in whole serum.

Returning to the results displayed in Figure 5 (*in vitro* selection cleavage results), another notable trend perceived in the overall selection process is the observation that counter selection generally produces higher cleavage levels than positive selection. This can be explained by multiple factors. Counter selection incubation time is longer than positive selection, which allows more time and opportunities for the DNAzyme library to interact with components of the serum. Additionally, the counter selection cancer pool is comprised of a larger number of samples both, which introduce more heterogeneity to the samples from individual to individual, and across multiple types of

cancer. This leads to the possibility of the counter selection cancer pool holding more targets for the library to interact with.

A third notable trend in the *in vitro* selection results is the steep increase in library cleavage following round 11 for both counter and positive selection. The change in incubation time likely accounts for a significant portion of this increased cleavage or can be a factor of prolonged exposure to nuclease activity in serum. The effect of nucleases on the DNAzyme library will be discussed in more detail later in this chapter. The decision to change the selection pressures in favour of more aggressive counter selection came from the consistent similarity between the percent cleavage in counter and positive selection for the first 10 rounds of selection. This finding was confirmed by sequencing analysis performed on the recovered DNAzyme libraries from Rounds 6-10, which demonstrated a lack of enrichment in the latter rounds of selection. Specifically, sequencing analysis showed that enrichment was occurring through Round 8, indicating that the selection pressures are adequate for performing the selection protocol, and that the library is becoming less diverse (determined by a decrease in the number of unique sequences in the library over the rounds of selection) in favour of sequences that can distinguish pancreatic cancer serum from counter selection serum. The sequencing results furthermore demonstrated a lack of enrichment in Rounds 9 and 10, which is an indication that we have exhausted the current selection conditions, and further enrichment requires altering the selection pressures.

Consequently, one promising option for increasing the selection pressures in favour of aggressive counter selection lies in the amendment of the counter selection incubation time. As such, we raised the incubation time of the DNAzyme library with the counter selection serum from 2 hours

to 24 hours, allowing us to filter out a larger number of DNAzyme sequences active in the combined control cancer serum in each round of selection, and effectively increasing the specificity of the emerging DNAzyme library. The positive selection incubation time was kept at 1 hour in order to maintain high affinity and sensitivity to pancreatic cancer serum.

Unsurprisingly, the new selective pressures led to a large jump in percent cleavage in Rounds 11 and 12, although it was once again consistent in both counter and positive selection. As such, an additional amendment was initiated for Round 13 and 14 to enable even more aggressive counter selection. The counter selection incubation time was further increased to 48 hours, while the positive selection incubation time was still held at 1 hour. This change resulted in a drastic increase in DNAzyme cleavage in the counter selection group, followed by a modest increase in the positive selection group. These results may represent an indication of the library beginning to lose its diversity, which would suggest that we are exhausting the library, leading to a natural stopping point for the *in vitro* selection protocol.

2.2.2 Identification of candidate DNAzyme sequences targeting pancreatic cancer

Sequencing Results

Rounds 6-10

Upon completion of the 10th round of selection, the recovered libraries from Rounds 6, 7, 8, 9, and 10 were sent for external sequencing with the Illumina MiSeq Next Generation Sequencing method. The results revealed a total of 746,464 classes of sequences in the Round 6 population, 800,927 in the Round 7 population, 669,257 in the Round 8 population, 279,379 in the Round 9

population, and 292,771 in the Round 10 population. The prevailing motifs for each round are depicted in Figure 8. A distinguished motif GTCTTAG was observed at the 30-35 nucleotide segment and the 40-45 nucleotide segment in the 7^{th} and 8^{th} round of selection, respectively. Interestingly, this motif has shown the same pattern of early emergence and disappearing in an unrelated selection for ALS currently underway in our research group, suggesting non-specific interactions with a common serum target. Ongoing enrichment was observed through Round 9, with a tapering thereafter.

From the insight given into the exhaustion of selection conditions at Round 10, we decided to perform an additional benchmark cleavage activity analysis with the top 3 ranking sequences by frequency from Round 10. Their activity in buffer, counter selection serum, and positive selection serum was compared to the top ranked sequence at Round 7 and the starting population (R0). This experiment serves to identify any early occurring sequences that already show preferential selectivity for pancreatic cancer serum, and as a measure of the finessing of enrichment from Rounds 7-10. We expect the top ranked sequence at Round 7 to perform worse than the Round 10 sequences as the library continues to evolve through the selection. This experiment consisted of ordering the top sequence at Round 7 and the three most common sequences at Round 10, ligating them with the FQ30 signal generating segment, and performing cleavage time course experiments with these DNAzymes in buffer, counter selection serum, and positive selection serum. The results of this experiment are illustrated in Figure 7.

A

	Round 7		Round 10	
Total Classes	800,927		292,771	
Sequences	**Rank**	**Frequency (%)**	**Rank**	**Frequency (%)**
7-3	3	0.00643	15	0.07435
10-1	1	0.01200	1	3.54872
10-2	2	0.00812	2	0.98828
10-3	11	0.00327	3	0.62839

B

7-3 5'-CTATGAACTGAC-Q-R-F-GACCTCACTACCAAGTTACGTCAAGGTGTCACTCCACGCGGACAATCTGTCTTAGTTAGTTCAAGTAGTGCGTGATCATTGTTTCAGTGGCGAGCA

10-1 5'-CTATGAACTGAC-Q-R-F-GACCTCACTACCAAGTTACGTCAAGGTGTCACTCCCAGCCATTGGTTGTATGGAGGTTCGATGTAAACTGTGAGGGTCATTGTTTCAGTGGCGAGCA

10-2 5'-CTATGAACTGAC-Q-R-F-GACCTCACTACCAAGTTACGTCAAGGTGTCACTCCCCATGCACGGTTTTGGACAAATAAGTGGGGTTGTTAGTCGTCATTGTTTCAGTGGCGAGCA

10-3 5'-CTATGAACTGAC-Q-R-F-GACCTCACTACCAAGTTACGTCAAGGTGTCACTCCCCTGGTGTCGACTAGTTCACTTGAGGTTGTTAGTCGGTATTCATTGTTTCAGTGGCGAGCA

C

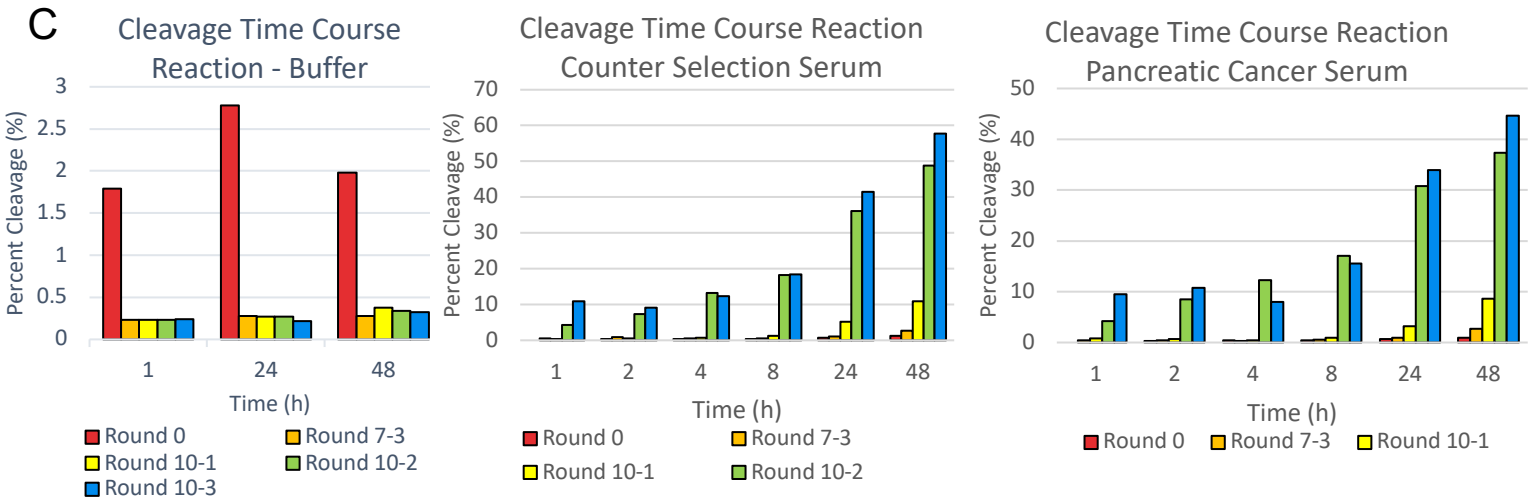


Figure 7. A. Sequencing results summary. **B.** Unique sequences of 7-3, 10-1, 10-2, and 10-3. **C.** Graph illustration of cleavage time course results.

Notably, minimal DNAzyme cleavage was observed in the buffer experiment. This finding has 2 major implications. First, it confirms that none of the 3 sequences is a self-cleaving sequence. Second, at 48 hours, the cleavage remained under 0.4%, indicating high internal DNAzyme stability at room temperature in solution. Another finding of interest is the significant increase in percent cleavage of the 10-2 and 10-3 sequences over time in both the control and pancreatic cancer sera. Looking specifically at 10-3, once again there appeared to be higher cleavage in the control serum than the pancreatic cancer serum, reaching a peak percent cleavage of 57.7% and 44.6%, respectively. This finding has further contributed to the decision to amend the protocol for subsequent rounds of selection, in order to increase the stringency of the counter selection portion of the experiment, and eliminate any sequences active in control serum.

Rounds 11-14

Upon completion of the 14th round of selection, the recovered libraries from Rounds 11, 12, 13, and 14 were sent for external sequencing. The results revealed a total of 204,341 classes of sequences in the Round 11 population, 136,759 in the Round 12 population, 75,541 in the Round 13 population, and 77,295 in the Round 14 population. The combined motifs from the first two sequencing rounds are combined in Figure 8. The dominant motif GTCTTAG present in Rounds 6-10 appears to disappear beginning at Round 8 and is replaced by a new dominant motif beginning at Round 11, namely GGTTGTTAG. While this finding appears promising, upon reviewing the raw sequence data from the 10th round of selection, the new motif was present in the 2nd and 3rd most common sequence in Round 10. Therefore, it has undergone individual testing with all 3 media (buffer, counter selection serum, and positive selection serum) without significant success, as will be detailed below. This finding, in combination with the decrease in enrichment as observed

by the plateauing number of classes of sequences, plus the high copy number of the top 50 sequences by Round 14 indicates that the library is reaching exhaustion, suggesting the conclusion of the selection.

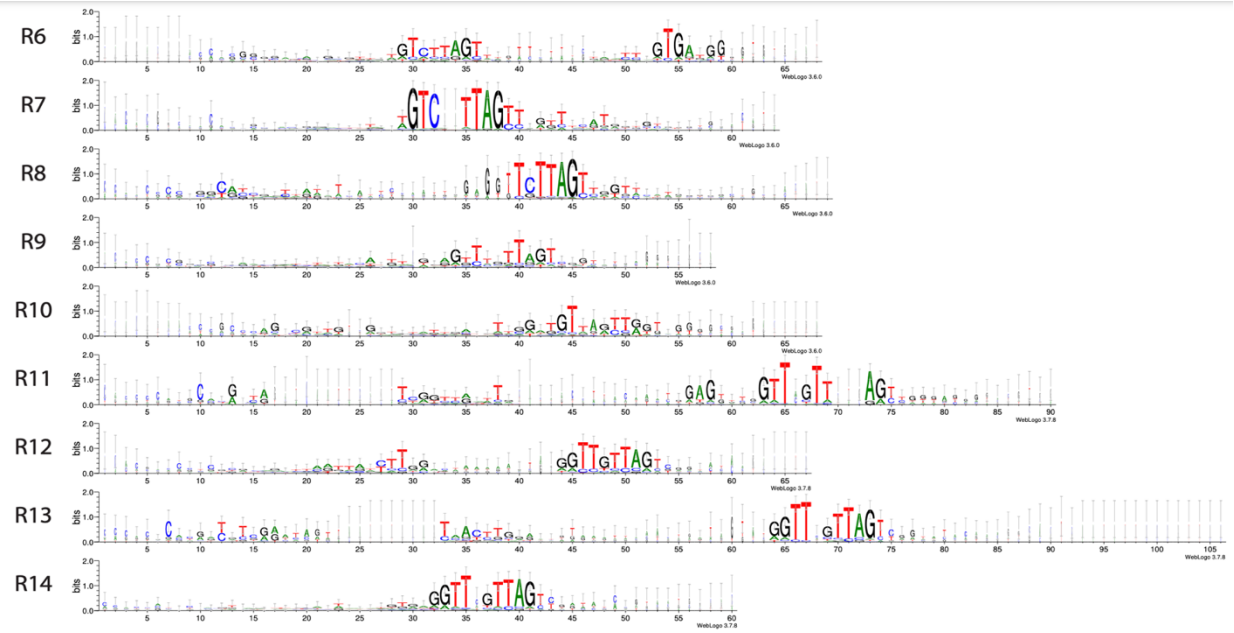


Figure 8. Round 6-14 sequencing results. Emerging motifs are formed by the frequency of the nucleotide found at each position in the random domain.

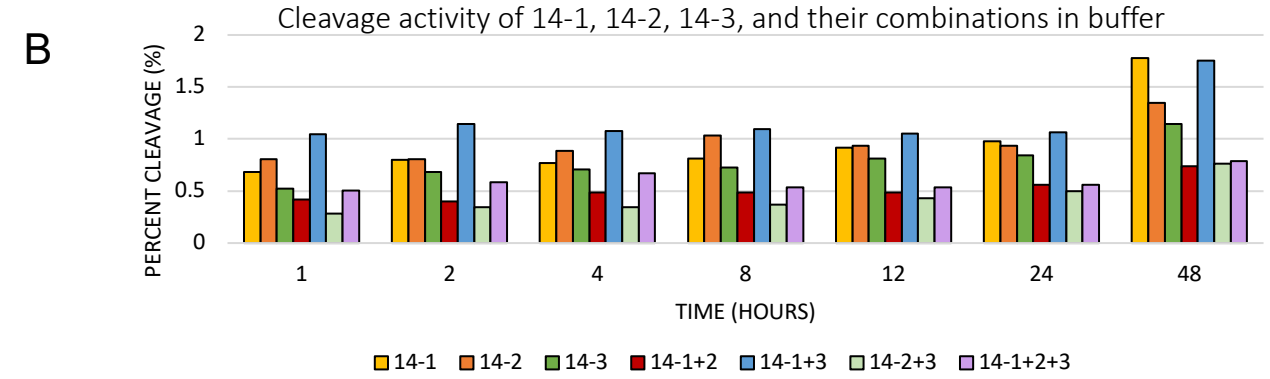
Despite having tested the top three sequences in Round 10, we opted to examine them again in more depth. Once again, each sequences underwent a cleavage time course reaction in buffer, control serum, and pancreatic cancer serum. Additionally, we opted to explore a new hypothesis with regards to DNAzyme performance in serum activity – namely the possibility that cleavage/responsiveness of one sequence may be enhanced by the presence of another sequence in solution. Whether this extra sequence forms a complex with the main DNAzyme to enhance function or simply facilitates the interaction between DNAzyme and target would be difficult to ascertain at this point. However, a worthwhile first step is to acknowledge whether the presence

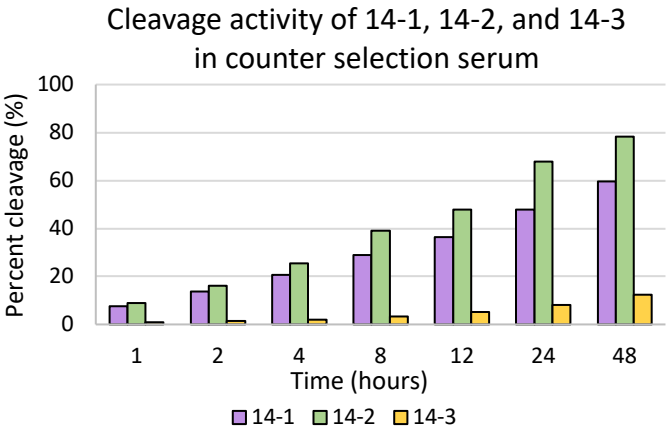
of another DNAzyme in solution carries any effect on the overall cleavage activity. Consequently, cleavage time course reactions were performed with combinations of the top three sequences at Round 14. With more time, this hypothesis would have also been tested with the addition of the entire Round 14 DNA library to each of the top three sequences at Round 14. The results of this experiment are profiled in Figure 9.

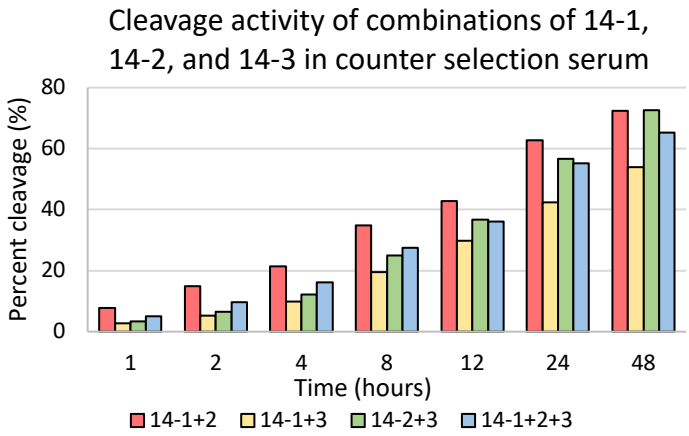
A

	Round 10		Round 14	
Total Classes	292,771		77,295	
Sequences	**Rank**	**Frequency (%)**	**Rank**	**Frequency (%)**
14-1	2	0.99	1	5.27
14-2	3	0.63	2	4.97
14-3	1	3.55	3	3.65

B







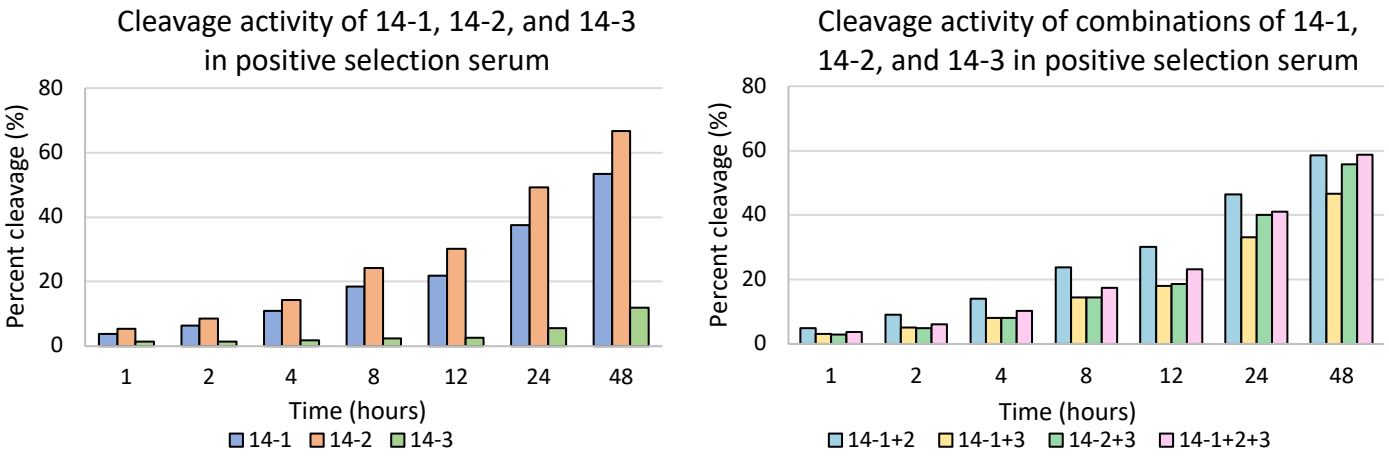


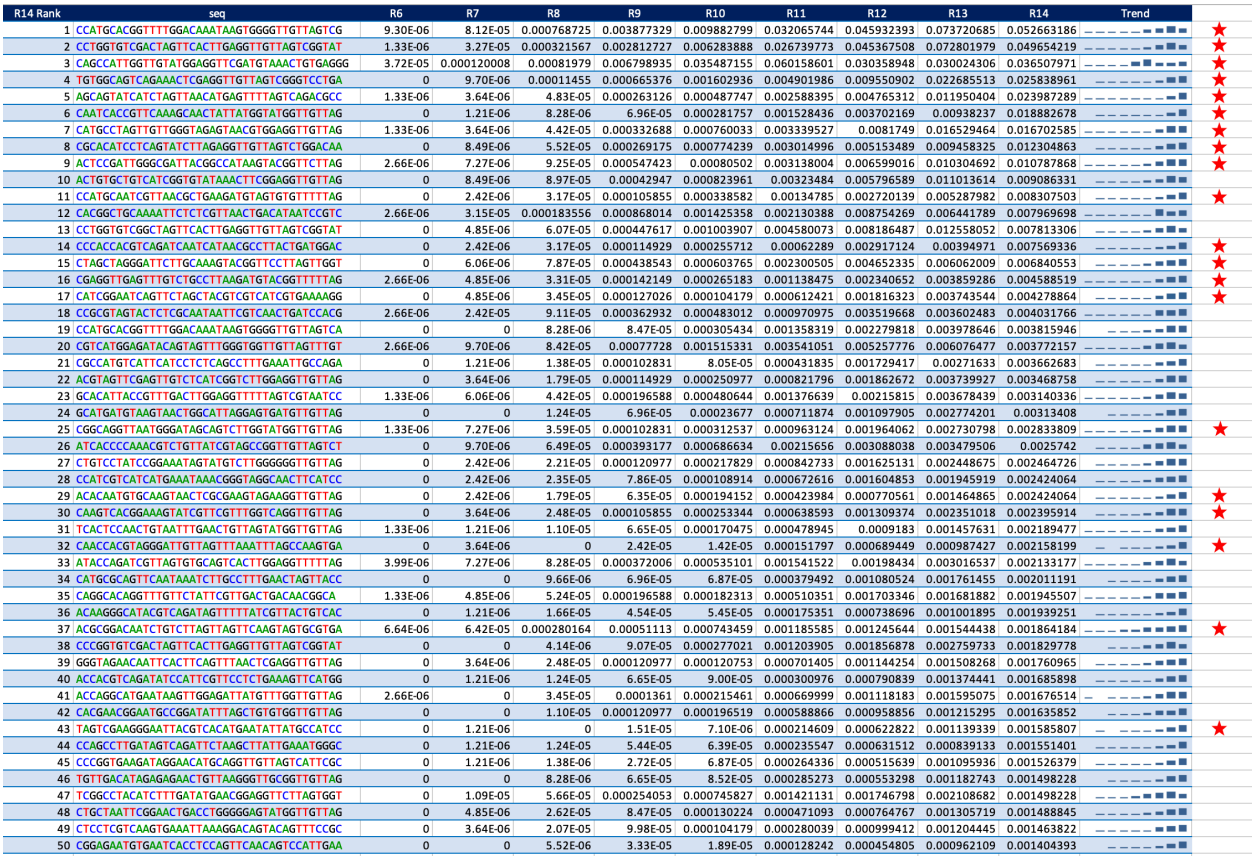
Figure 9. A. Sequencing information from Round 10 and 14. **B.** Graph representation of cleavage percentages of top 3 sequences at Round 14.

Sequencing data from Figure 9 shows a reshuffling of the order of the top sequences from Round 10 to 14, as well as an overall increase in the frequency/copy number of each sequence in the library. At first glance, the persistence of these sequences through four rounds of selection suggests a strong association with serum components; however, upon analysing their frequency trends from Rounds 10 to 14, they appear to be on the decline, having peaked by Round 13. This indicates that they are being phased out of the enriched library, coinciding with the introduction of the more stringent counter selection pressures.

We now turn our attention to the cleavage time course reactions. Beginning with the buffer medium, all sequences show low self-cleavage and acceptable functional stability through 48 hours. 14-1 appears to have the highest amount of self-cleavage/degradation. The raw cleavage data also suggests a potentially deleterious interaction between 14-1 and 14-3, since their combination produced higher cleavage values than either individual sequences. This observation

must be taken in the context of overall very low cleavage values, and the effect (if any) is very small and within margin of error. Next, the counter and positive selection serum data unfortunately shows that all three sequences cleave preferentially in counter selection serum, although the raw cleavage values in both the positive and counter selection serum are quite staggering, and approximate each other in values. This does show that the selection objectively was successful, in that we have produced sequences that are much more readily able to detect components of human cancer compared to the starting library. Across both positive and counter selection, 14-2 performed the best out of the three sequences, reaching values of 66.7% and 78.4%, respectively at 48h. Lastly, there appears to be little to no effect of combining sequences on the overall cleavage percentage of each time course reaction, as the combined values roughly reflect the average of the cleavage percentages of the individual sequence components.

Given the lack of specificity of the top three sequences at Round 14, it is unsurprising to see their frequency on the decline. Consequently, we opted to have a deeper look at the sequencing data from Round 14, in order to select additional candidates for sensitivity and specificity testing. The top fifty sequences from Round 14 are illustrated in Figure 9A. After reviewing the trends in the frequency for each of the candidates through the prior 9 rounds of selection, we compiled a list of contending sequences that merit further sensitivity and specificity evaluation (marked by a star in Figure 10.



R14 Rank	seq	R6	R7	R8	R9	R10	R11	R12	R13	R14	Trend	
1	CCATGCACGGTTTTGGACAAATAAGTGGGGTTGTTAGTCG	9.30E-06	8.12E-05	0.000768725	0.003877329	0.009882799	0.032065744	0.045932393	0.073720685	0.052663186		★
2	CCTGGTGTCGACTAGTTCACTTGAGGTTGTTAGTCGGTAT	1.33E-06	3.27E-05	0.000321567	0.002812727	0.006283888	0.026739773	0.045367508	0.072801979	0.049654219		★
3	CAGCCATTGGTTGTATGGAGGTTCGATGTAAACTGTGAGGG	3.72E-05	0.000120008	0.00081979	0.006798935	0.035487155	0.060158601	0.030358948	0.030024306	0.036507971		★
4	TGTGGCAGTCAGAAACTCGAGGTTGTTAGTCGGGTCCTGA	0	9.70E-06	0.00011455	0.000665376	0.001602936	0.004901986	0.009550902	0.022685513	0.025838961		★
5	AGCAGTATCATCTAGTTAACATGAGTTTTAGTCAGACGCC	1.33E-06	3.64E-06	4.83E-05	0.000263126	0.000487747	0.002588395	0.004765312	0.011950404	0.023987289		★
6	CAATCACCGTTCAAAGCAACTATTATGGTATGGTTGTTAG	0	1.21E-06	8.28E-06	6.96E-05	0.000281757	0.001528436	0.003702169	0.00938237	0.018882678		★
7	CATGCCTAGTTGTTGGGTAGAGTAACGTGGAGGTTGTTAG	1.33E-06	3.64E-06	4.42E-05	0.000332688	0.000760033	0.003339527	0.0081749	0.016529464	0.016702585		★
8	CGCACATCCTCAGTATCTTAGAGGTTGTTAGTCTGGACAA	0	8.49E-06	5.52E-05	0.000269175	0.000774239	0.003014996	0.005153489	0.009458325	0.012304863		★
9	ACTCCGATTGGGCGATTACGGCCATAAGTACGGTTCTTAG	2.66E-06	7.27E-06	9.25E-05	0.000547423	0.00080502	0.003138004	0.006599016	0.010304692	0.010787868		★
10	ACTGTGCTGTCATCGGTGTATAAACTTCGGAGGTTGTTAG	0	8.49E-06	8.97E-05	0.00042947	0.000823961	0.00323484	0.005796589	0.011013614	0.009086331		
11	CCATGCAATCGTTAACGCTGAAGATGTAGTGTGTTTTTAG	0	2.42E-06	3.17E-05	0.000105855	0.000338582	0.00134785	0.002720139	0.005287982	0.008307503		★
12	CACGGCTGCAAAATTCTCTCGTTAACTGACATAATCCGTC	2.66E-06	3.15E-05	0.000183556	0.000868014	0.001425358	0.002130388	0.008754269	0.006441789	0.007969698		
13	CCTGGTGTCGGCTAGTTCACTTGAGGTTGTTAGTCGGTAT	0	4.85E-06	6.07E-05	0.000447617	0.001003907	0.004580073	0.008186487	0.012558052	0.007813306		
14	CCCACCACGTCAGATCAATCATAACGCCTTACTGATGGAC	0	2.42E-06	3.17E-05	0.000114929	0.000255712	0.00062289	0.002917124	0.00394971	0.007569336		★
15	CTAGCTAGGGATTCTTGCAAAGTACGGTTCCTTAGTTGGT	0	6.06E-06	7.87E-05	0.000438543	0.000603765	0.002300505	0.004652335	0.006062009	0.006840553		★
16	CGAGGTTGAGTTTGTCTGCCTTAAGATGTACGGTTTTTAG	2.66E-06	4.85E-06	3.31E-05	0.000142149	0.000265183	0.001138475	0.002340652	0.003859286	0.004588519		★
17	CATCGGAATCAGTTCTAGCTACGTCGTCATCGTGAAAAGG	0	4.85E-06	3.45E-05	0.000127026	0.000104179	0.000612421	0.001816323	0.003743544	0.004278864		★
18	CCGCGTAGTACTCTCGCAATAATTCGTCAACTGATCCACG	2.66E-06	2.42E-05	9.11E-05	0.000362932	0.000483012	0.000970975	0.003519668	0.003602483	0.004031766		
19	CCATGCACGGTTTTGGACAAATAAGTGGGGTTGTTAGTCA	0	0	8.28E-06	8.47E-05	0.000305434	0.001358319	0.002279818	0.003978646	0.003815946		
20	CGTCATGGAGATACAGTAGTTTGGGTGGTTGTTAGTTTGT	2.66E-06	9.70E-06	8.42E-05	0.00077728	0.001515331	0.003541051	0.005257776	0.006076477	0.003772157		
21	CGCCATGTCATTCATCCTCTCAGCCTTTGAAATTGCCAGA	0	1.21E-06	1.38E-05	0.000102831	8.05E-05	0.000431835	0.001729417	0.00271633	0.003662683		
22	ACGTAGTTCGAGTTGTCTCATCGGTCTTGGAGGTTGTTAG	0	3.64E-06	1.79E-05	0.000114929	0.000250977	0.000821796	0.001862672	0.003739927	0.003468758		
23	GCACATTACCGTTTGACTTGGAGGTTTTTAGTCGTAATCC	1.33E-06	6.06E-06	4.42E-05	0.000196588	0.000480644	0.001376639	0.00215815	0.003678439	0.003140336		
24	GCATGATGTAAGTAACTGGCATTAGGAGTGATGTTGTTAG	0	0	1.24E-05	6.96E-05	0.00023677	0.000711874	0.001097905	0.002774201	0.00313408		
25	CGGCAGGTTAATGGGATAGCAGTCTTGGTATGGTTGTTAG	1.33E-06	7.27E-06	3.59E-05	0.000102831	0.000312537	0.000963124	0.001964062	0.002730798	0.002833809		★
26	ATCACCCCAAACGTCTGTTATCGTAGCCGGTTGTTAGTCT	0	9.70E-06	6.49E-05	0.000393177	0.000686634	0.00215656	0.003088038	0.003479506	0.0025742		
27	CTGTCCTATCCGGAAATAGTATGTCTTGGGGGGTTGTTAG	0	2.42E-06	2.21E-05	0.000120977	0.000217829	0.000842733	0.001625131	0.002448675	0.002464726		
28	CCATCGTCATCATGAAATAAACGGGTAGGCAACTTCATCC	0	2.42E-06	2.35E-05	7.86E-05	0.000108914	0.000672616	0.001604853	0.001945919	0.002424064		
29	ACACAATGTGCAAGTAACTCGCGAAGTAGAAGGTTGTTAG	0	2.42E-06	1.79E-05	6.35E-05	0.000194152	0.000423984	0.000770561	0.001464865	0.002424064		★
30	CAAGTCACGGAAAGTATCGTTCGTTTGGTCAGGTTGTTAG	0	3.64E-06	2.48E-05	0.000105855	0.000253344	0.000638593	0.001309374	0.002351018	0.002395914		★
31	TCACTCCAACTGTAATTTGAACTGTTAGTATGGTTGTTAG	1.33E-06	1.21E-06	1.10E-05	6.65E-05	0.000170475	0.000478945	0.0009183	0.001457631	0.002189477		
32	CAACCACGTAGGGATTGTTAGTTTAAATTTAGCCAAGTGA	0	3.64E-06	0	2.42E-05	1.42E-05	0.000151797	0.000689449	0.000987427	0.002158199		★
33	ATACCAGATCGTTAGTGTGCAGTCACTTGGAGGTTTTTAG	3.99E-06	7.27E-06	8.28E-05	0.000372006	0.000535101	0.001541522	0.00198434	0.003016537	0.002133177		
34	CATGCGCAGTTCAATAAATCTTGCCTTTGAACTAGTTACC	0	0	9.66E-06	6.96E-05	6.87E-05	0.000379492	0.001080524	0.001761455	0.002011191		
35	CAGGCACAGGTTTGTTCTATTCGTTGACTGACAACGGCA	1.33E-06	4.85E-06	5.24E-05	0.000196588	0.000182313	0.000510351	0.001703346	0.001681882	0.001945507		
36	ACAAGGGCATACGTCAGATAGTTTTTATCGTTACTGTCAC	0	1.21E-06	1.66E-05	4.54E-05	5.45E-05	0.000175351	0.000738696	0.001001895	0.001939251		
37	ACGCGGACAATCTGTCTTAGTTAGTTCAAGTAGTGCGTGA	6.64E-06	6.42E-05	0.000280164	0.00051113	0.000743459	0.001185585	0.001245644	0.001544438	0.001864184		★
38	CCCGGTGTCGACTAGTTCACTTGAGGTTGTTAGTCGGTAT	0	0	4.14E-06	9.07E-05	0.000277021	0.001203905	0.001856878	0.002759733	0.001829778		
39	GGGTAGAACAATTCACTTCAGTTTAACTCGAGGTTGTTAG	0	3.64E-06	2.48E-05	0.000120977	0.000120753	0.000701405	0.001144254	0.001508268	0.001760965		
40	ACCACGTCAGATATCCATTCGTTCCTCTGAAAGTTCATGG	0	1.21E-06	1.24E-05	6.65E-05	9.00E-05	0.000300976	0.000790839	0.001374441	0.001685898		
41	ACCAGGCATGAATAAGTTGGAGATTATGTTTGGTTGTTAG	2.66E-06	0	3.45E-05	0.0001361	0.000215461	0.000669999	0.001118183	0.001595075	0.001676514		
42	CACGAACGGAATGCCGGATATTTAGCTGTGTGGTTGTTAG	0	0	1.10E-05	0.000120977	0.000196519	0.000588866	0.000958856	0.001215295	0.001635852		
43	TAGTCGAAGGGAATTACGTCACATGAATATTATGCCATCC	0	1.21E-06	0	1.51E-05	7.10E-06	0.000214609	0.000622822	0.001139339	0.001585807		★
44	CCAGCCTTGATAGTCAGATTCTAAGCTTATTGAAATGGGC	0	1.21E-06	1.24E-05	5.44E-05	6.39E-05	0.000235547	0.000631512	0.000839133	0.001551401		
45	CCCGGTGAAGATAGGAACATGCAGGTTGTTAGTCATTCGC	0	1.21E-06	1.38E-06	2.72E-05	6.87E-05	0.000264336	0.000515639	0.001095936	0.001526379		
46	TGTTGACATAGAGAGAACTGTTAAGGGTTGCGGTTGTTAG	0	0	8.28E-06	6.65E-05	8.52E-05	0.000285273	0.000553298	0.001182743	0.001498228		
47	TCGGCCTACATCTTTGATATGAACGGAGGTTCTTAGTGGT	0	1.09E-05	5.66E-05	0.000254053	0.000745827	0.001421131	0.001746798	0.002108682	0.001498228		
48	CTGCTAATTCGGAACTGACCTGGGGGAGTATGGTTGTTAG	0	4.85E-06	2.62E-05	8.47E-05	0.000130224	0.000471093	0.000764767	0.001305719	0.001488845		
49	CTCCTCGTCAAGTGAAATTAAAGGACAGTACAGTTTCCGC	0	3.64E-06	2.07E-05	9.98E-05	0.000104179	0.000280039	0.000999412	0.001204445	0.001463822		
50	CGGAGAATGTGAATCACCTCCAGTTCAACAGTCCATTGAA	0	0	5.52E-06	3.33E-05	1.89E-05	0.000128242	0.000454805	0.000962109	0.001404393		

Figure 10. Round 14 sequencing results showing the random domains of the top 50 sequences by frequency (fraction of total library). The "Trend" column demonstrates their frequency through the prior 9 rounds of selection. Candidate sequences selected for sensitivity and specificity testing are marked by a red star.

2.2.3 Sensitivity and Specificity analysis

The first step in the analysis of the sensitivity and specificity of the candidate sequences involves testing their performance in selection buffer, to ensure no self-cleaving sequences have been selected. Figure 10 illustrates the results of the cleavage time course reactions of the 20 candidate sequences in buffer. All sequences demonstrate minimal background cleavage below 1% through 24h, indicating they are all viable candidates to proceed with sensitivity and specificity analysis.

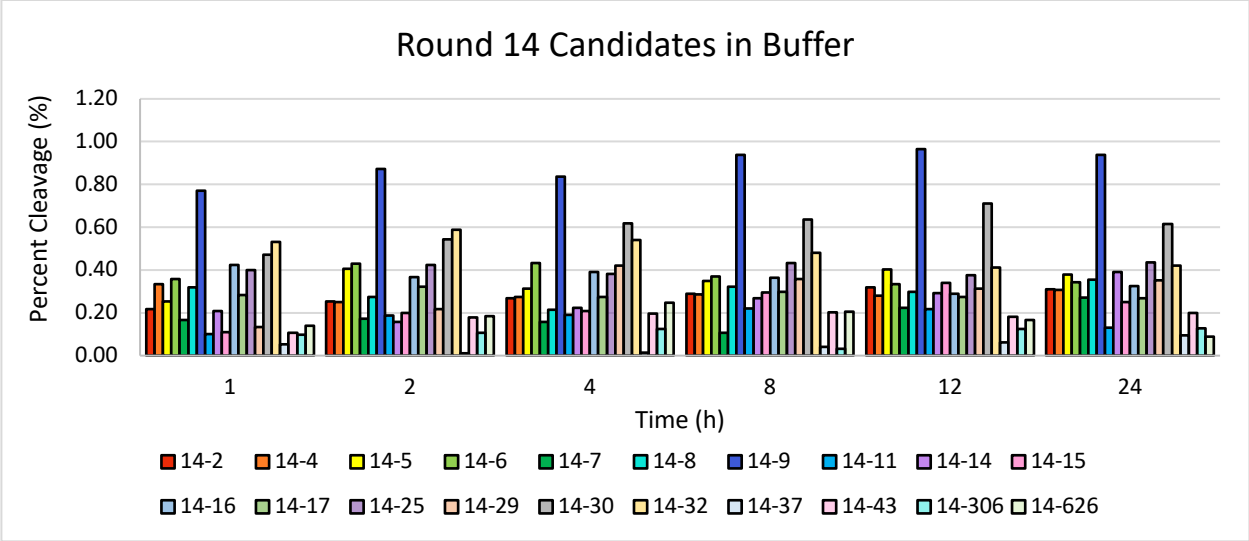
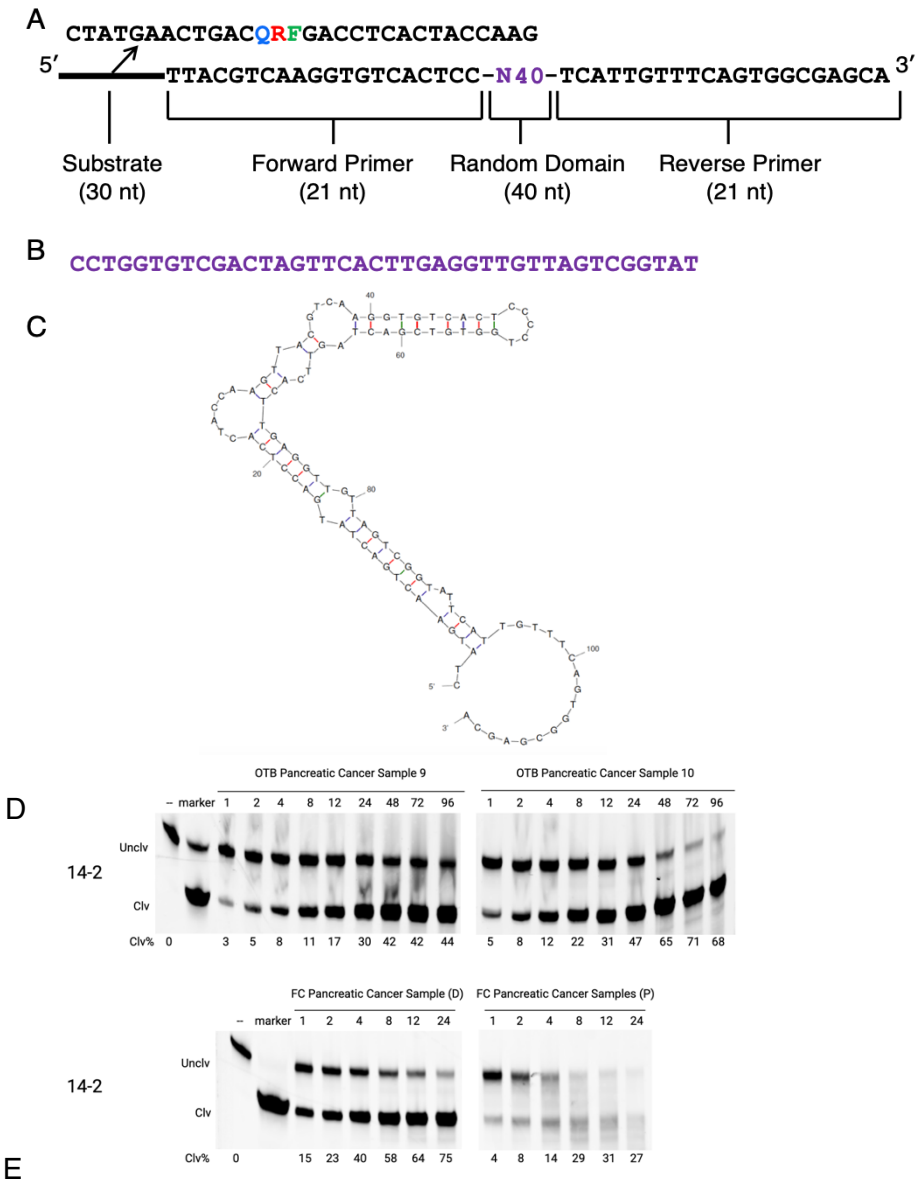


Figure 11. Round 14 candidates in selection buffer CTCRs.

Sensitivity

Given the previously demonstrated high cleavage values of 14-2, we opted to include it despite its low specificity in pooled control serum and explore its performance in individual patient samples. Testing 14-2 in individual patient samples serves to conclude (1) whether its high sensitivity in pooled pancreatic cancer serum is sustained in individual samples, and (2) whether the effect of the pooled control serum had any compounding effect on the cleavage patterns of 14-2. Figure 12 shows the construction of the 14-2 DNAzyme, its individual sequence, predicted folding structure, and its performance in individual pancreatic cancer samples.



Time (h)	PancPool	Panc2	Panc3	Panc4	Panc5	Panc6	Panc7	Panc8	Panc9	Panc10	Panc11	Panc12	Panc13	Panc14	Panc15
1	6.11	4.85	3.93	7.23	5.71	3.35	5.64	8.62	5.21	2.73	4.73	3.81	4.65	6.86	7.66
2	11.05	9.27	5.89	12.68	10.60	6.35	9.38	14.85	8.72	5.27	7.94	6.11	7.43	10.14	12.57
4	15.57	15.67	9.96	19.91	18.20	9.66	17.69	22.64	14.64	7.69	12.47	9.78	11.96	16.34	22.13
8	26.24	24.34	16.76	29.97	29.50	15.27	27.76	31.83	23.94	11.46	22.49	17.19	20.31	30.82	36.05
12	32.74	31.75	21.82	43.37	41.21	21.13	34.63	45.76	30.65	17.05	31.47	22.85	24.85	36.35	40.40
24	47.87	48.96	35.24	59.97	58.40	35.62	48.05	54.30	47.35	30.01	46.88	37.61	35.99	30.62	37.58
48	64.15	67.92	53.78	74.68	70.73	49.67	57.27	61.80	63.60	42.30	64.90	37.98	39.43	22.97	29.43
72	65.92	68.75	53.84	72.64	72.98	57.26	66.48	59.13	65.86	42.25	70.68	30.87	37.89	24.38	28.45
96	70.29	73.21	56.15	83.49	75.64	64.28	69.17	69.38	68.93	44.01	68.19	28.22	35.20	17.69	19.67

Figure 12. Sequence of 14-2 DNAzyme. A. Ligation of PanC2 library with FQ30 segment. B. Random domain sequence for 14-2. C. Predicted folding structure of 14-2 DNAzyme. D. Performance of 14-2 in four individual patient samples of pancreatic cancer. E. Graphical representation of the cleavage values in each pancreatic cancer patient sample cleavage time course. 2-10 = OTB samples 11=FcDefib (Fresh collected defibrinated) 12-15= FCSerum (Fresh collected serum)

Figure 12D denotes consistent cleavage activity of 14-2 in both the OTB pancreatic cancer samples involved in its selection, and in the fresh collected local area patient samples. This DNAzyme activity comparison is the first step in demonstrating validation and indicating potential for reproducibility and generalized sensitivity of 14-2 towards pancreatic cancer detection. It should be noted that the fresh collected samples depicted in this comparison have two distinct modes of preparation. The sample on the left of the figure marked (D) was prepared identically to the OTB samples – namely converted from whole blood to plasma and subsequently defibrinated. Consequently, the most salient difference between the two samples is their collection age. The fresh defibrinated sample bears strikingly similar cleavage values to the OTB samples, an early sign of generalized sensitivity of 14-2 towards pancreatic cancer detection.

The fresh collected sample on the right of Figure 12D is a pool of pancreatic cancer patient samples directly seroconverted from whole blood. In this fresh serum preparation, 14-2 shows marked degradation compared to the OTB samples and the defibrinated fresh sample. This is evidenced by the uncleaved band dissipating significantly after 4 hours compared to the OTB samples and the fresh defibrinated sample, without a converse increase in fluorescence emission of the cleaved band. Combined with the observation of multiple small fluorescent cleavage fragments downstream from the cleavage band, and the eventual decrease in cleavage after a peak at 12 hours, this paints the picture of non-specific serum endonuclease degradation of the DNAzyme probe.

Although not as pronounced, there does appear to be a perceptible amount of degradation of the DNAzyme probe over time in the OTB and fresh defibrinated sample as well, particularly after 24 hours. Using OTB Pancreatic Cancer Sample 10 as an example, there is noted dissipation of the

uncleaved band after 24 hours. Upon reviewing the raw fluorescence data, this does not correlate with an increase in the fluorescence emission of the cleaved band. Additionally, truncated/digested fluorescent fragments of the DNAzyme were visible downstream from the cleavage band in this sample as well, coinciding with the onset of uncleaved band dissipation at 24 hours. Lastly, a DNAzyme cleavage peak is once again observed prior to the end of the time course, although delayed until 72 hours compared to the fresh collected pancreatic serum pool. An exploration of the different rates of degradation based on serum sample preparation and contributing causes will be explored in Chapter 5 of this thesis.

Figure 12E is a compilation of the performance of 14-2 in 14 individual pancreatic cancer patient samples (OTB and fresh collected), and one pooled sample comprised of the OTB pancreatic cancer patient samples. Of note, samples labelled Panc2-11 are the OTB samples involved in the selection of 14-2. Samples labelled Panc12-14 represent the fresh collected pancreatic cancer patient samples directly seroconverted from whole blood. The fresh collected defibrinated sample was not available for inclusion in this experiment.

The purpose of including the pooled sample was to test whether the DNAzyme's performance in pooled serum approximates its performance in individual patient samples, which was readily demonstrated. Additionally, it appears the 14-2 DNAzyme is more stable and more responsive to the OTB samples than the fresh collected samples, as evidenced by higher rates of cleavage at 96 hours and by comparing peak cleavage values. However, this observation comes with the caveat of the DNAzyme undergoing much more non-specific degradation of both the uncleaved and

cleaved form of the DNAzyme in the fresh samples, as demonstrated by early peaks of cleavage and subsequent decline in the fresh serum samples.

Three important takeaways can be obtained from the analysis of the 14-2 DNAzyme in individual patient samples. First, fresh collected serum samples contain higher amounts of endonucleases or more active endonucleases compared to stored serum samples. This observation is particularly observed in samples directly converted from whole blood to serum. The second conclusion from this analysis points to the length of time courses. Specifically, regardless of the sample origin/preparation, 96-hour time courses do not offer reliable additional information about cleavage patterns beyond 24 hours due to significant non-specific nuclease degradation of the DNAzyme. The third conclusion that can be drawn from this analysis refers to the behaviour of 14-2 in pooled serum samples compared to individual samples. In fact, the activity of 14-2 in the pancreatic cancer serum sample pool appears to reflect the activity of 14-2 in the pool's constituents, which is an advantageous finding with regards to expediting the remaining analysis of candidate sequences. Consequently, when planning the sensitivity screening of the remaining candidates, we opted to limit the time courses to 24 hours, and performed them in pooled patient samples. Figure 13 showcases the Round 14 candidate DNAzymes' performance in pooled OTB cancer serum.

Pancreatic - OTB

Time (h)	R14-2	R14-4	R14-5	R14-6	R14-7	R14-8	R14-9	R14-11	R14-14	R14-15	R14-16	R14-17	R14-25	R14-29	R14-30	R14-32	R14-37	R14-43	R14-306	R14-626
1.5	4.69	0.27	0.28	0.23	0.05	0.20	0.58	0.35	0.54	0.51	0.08	0.61	0.47	0.21	0.41	0.83	0.15	1.58	0.17	0.15
3	8.38	0.23	0.17	0.24	0.05	0.23	0.63	0.37	0.95	0.42	0.07	1.23	0.61	0.20	0.51	1.49	0.15	2.73	0.11	0.06
4.5	14.33	0.15	0.28	0.36	0.11	0.17	0.57	0.36	1.49	0.48	0.19	2.46	0.64	0.29	0.35	1.88	0.20	5.17	0.20	0.06
6	27.50	0.21	0.28	0.26	0.33	0.11	0.66	0.37	3.15	0.52	0.20	4.21	0.86	0.28	0.39	3.39	0.20	9.31	0.26	0.24
14	34.02	0.31	0.38	0.60	0.29	0.20	0.55	0.48	4.09	0.61	0.13	6.16	0.72	0.30	0.39	4.88	0.17	14.42	0.46	0.34
24	55.99	0.40	0.51	0.46	0.45	0.37	0.72	0.50	9.92	0.80	0.28	16.27	1.15	0.41	0.47	9.72	0.45	34.71	0.60	0.42

Pancreatic - Fresh collected

Time (h)	R14-2	R14-4	R14-5	R14-6	R14-7	R14-8	R14-9	R14-11	R14-14	R14-15	R14-16	R14-17	R14-25	R14-29	R14-30	R14-32	R14-37	R14-43	R14-306	R14-626
1	3.92	0.74	1.43	0.32	0.64	0.35	0.66	0.36	1.23	0.83	0.70	0.55	0.54	0.92	0.57	0.97	0.49	0.96	0.21	0.66
2	7.55	0.93	2.08	0.66	0.45	0.43	0.84	0.39	1.68	1.17	0.92	0.80	0.74	1.23	0.65	1.37	0.69	1.71	0.35	0.81
4	14.32	2.16	2.56	0.54	0.92	0.77	0.95	0.84	2.64	2.28	0.94	1.10	1.13	1.95	0.94	2.05	1.01	3.74	0.53	1.12
8	28.57	3.54	3.30	0.87	1.16	1.07	0.92	1.31	3.78	4.91	1.36	1.66	2.47	3.09	1.39	3.80	1.51	6.60	1.10	1.39
12	30.75	33.87	5.14	2.09	2.79	2.48	1.80	1.76	4.10	7.51	1.99	2.44	3.31	4.93	1.84	5.87	2.05	8.02	1.29	2.48
24	26.81	37.28	6.14	3.41	5.93	8.02	2.79	2.96	8.21	16.61	4.45	5.60	10.82	8.01	3.84	8.02	3.46	10.69	2.50	6.13

Figure 13. Round 14 candidates' performance in pooled pancreatic cancer serum form OTB samples and fresh collected serum samples.

Figure 13A confirms that 14-2 holds the best sensitivity profile among the Round 14 DNAzyme candidates. However, additional notable prospects are 14-43, 14-17, 14-14, and 14-32, in order of their peak cleavage values at 24 hours. 14-43 was a late addition to the group of candidates, as a results of advanced bioinformatic analysis. Consequently, its selection as a candidate and performance will be discussed further in Chapter 4. Lastly and impressively, all candidates appear to peak at 24 hours in activity, with low amounts of non-specific degradation. This finding is confirmed in pooled fresh collected serum (Figure 13B), indicating that 24 hours seems to be a better cleavage time course incubation time.

The results in Figure 13B point to 14-4 having the best activity in pooled fresh collected pancreatic cancer serum, followed by 14-2. However, when compared to 14-4's activity in OTB samples, this rise in activity can be largely attributed to nuclease degradation. 14-43 continues to stand out following testing with pooled fresh collected serum samples, however not as strongly as in the analysis with OTB samples (Figure 13A). While 14-17 lags in activity in fresh serum compared to OTB serum, 14-14 and 14-32 have comparable values of cleavage in OTB serum and fresh

collected serum. Therefore, the most notable candidates to emerge from the sensitivity analysis are 14-2, 14-43, 14-14, and 14-32. The random domains of these candidates were explored along with 14-17, however no emerging motif or pattern was noted linking the sequences. Their sequences and structures are illustrated in Figure 14.

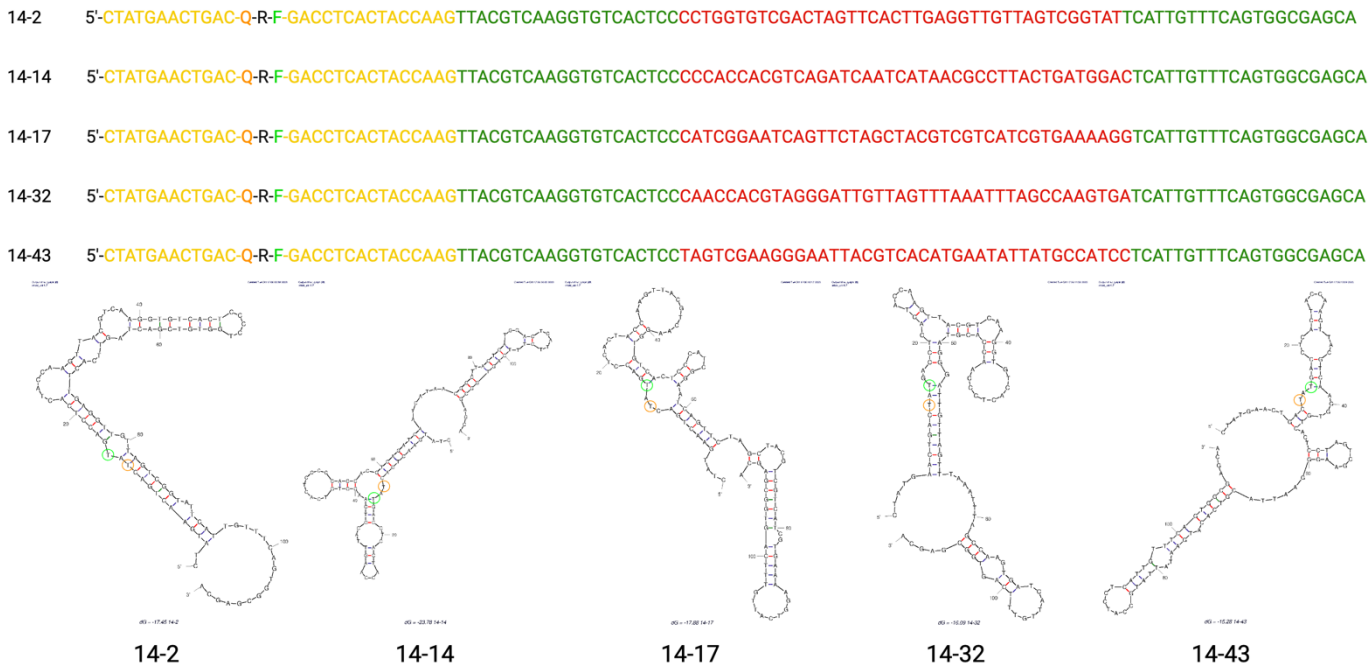


Figure 14. Sequences of contending DNAzymes emerging from the sensitivity analysis and their predicted folding structures. The fluorophore (green) and quencher (orange) locations along the sequences are encircled.

Specificity

Pursuant to the conclusions of the sensitivity analysis, specificity analyses on all 20 candidate sequences were conducted as 24-hour time courses, in pooled serum. Specifically, each sequence was tested in 4 pooled OTB cancer types (Breast, Colorectal, Lung, Prostate), and one fresh collected Healthy Participant serum pool. Each sample in the fresh collected healthy pool was

directly converted to serum from whole blood. The results of this broad specificity analysis are illustrated in Figure 15.

Breast

Time (h)	R14-2	R14-4	R14-5	R14-6	R14-7	R14-8	R14-9	R14-11	R14-14	R14-15	R14-16	R14-17	R14-25	R14-29	R14-30	R14-32	R14-37	R14-43	R14-306	R14-626
1	7.81	0.28	0.23	0.34	0.23	0.25	0.63	0.41	0.87	0.43	0.20	1.22	0.69	0.36	0.42	1.03	0.53	2.12	0.31	0.34
2	13.62	0.29	0.30	0.28	0.32	0.37	0.58	0.46	1.22	0.45	0.35	1.23	0.65	0.40	0.54	1.91	0.36	3.91	0.39	0.28
4	20.57	0.22	0.33	0.26	0.31	0.34	0.59	0.48	1.89	0.53	0.36	3.13	0.71	0.48	0.48	3.12	0.34	8.12	0.38	0.28
8	34.60	0.26	0.35	0.31	0.44	0.38	0.63	0.52	2.64	0.70	0.47	4.24	0.62	0.50	0.48	5.98	0.38	13.89	0.46	0.32
12	43.64	0.33	0.44	0.37	0.54	0.49	0.56	0.59	4.04	0.83	0.45	5.81	0.97	0.56	0.55	7.40	0.41	21.89	0.60	0.37
24	63.79	0.47	0.55	0.44	0.72	0.67	0.86	0.77	9.01	0.95	0.57	15.04	1.08	0.75	0.74	17.10	0.52	45.48	0.67	0.43

Colorectal

Time (h)	R14-2	R14-4	R14-5	R14-6	R14-7	R14-8	R14-9	R14-11	R14-14	R14-15	R14-16	R14-17	R14-25	R14-29	R14-30	R14-32	R14-37	R14-43	R14-306	R14-626
1	5.89	0.26	0.37	0.02	0.16	0.03	0.13	0.26	0.75	0.18	0.07	0.88	0.37	0.33	0.23	0.94	0.20	2.11	0.23	0.09
2	10.99	0.29	0.33	0.06	0.18	0.05	0.22	0.27	1.36	0.21	0.00	1.41	0.31	0.38	0.31	1.69	0.27	4.15	0.31	0.10
4	21.75	0.24	0.41	0.11	0.27	0.02	0.20	0.37	1.86	0.25	0.00	2.61	0.45	0.28	0.28	2.79	0.26	7.47	0.36	0.07
8	36.46	0.23	0.31	0.09	0.27	0.09	0.15	0.43	2.94	0.34	0.04	5.40	0.54	0.37	0.30	5.47	0.42	13.48	0.48	0.19
12	47.17	0.32	0.42	0.12	0.29	0.15	0.19	0.47	3.99	0.37	0.03	6.93	0.42	0.42	0.40	7.54	0.38	21.18	0.53	0.25
24	66.87	0.51	0.46	0.18	0.32	0.23	0.29	0.36	9.33	0.47	0.14	17.24	0.75	0.61	0.37	16.73	0.49	45.90	0.71	0.32

Lung

Time (h)	R14-2	R14-4	R14-5	R14-6	R14-7	R14-8	R14-9	R14-11	R14-14	R14-15	R14-16	R14-17	R14-25	R14-29	R14-30	R14-32	R14-37	R14-43	R14-306	R14-626
1	5.60	0.32	0.23	0.09	0.63	0.02	0.30	0.23	0.92	0.21	0.00	0.91	1.51	0.38	0.33	1.04	0.55	1.99	0.50	0.23
2	12.16	0.35	0.36	0.10	0.37	0.08	0.28	0.30	1.09	0.18	0.03	1.43	1.74	0.50	0.34	2.05	0.47	3.68	0.43	0.17
4	20.40	0.34	0.25	0.17	0.23	0.03	0.32	0.27	1.81	0.24	0.04	2.62	1.65	0.38	0.37	3.29	0.60	7.28	0.51	0.31
8	33.60	0.36	0.33	0.18	0.32	0.09	0.37	0.32	2.85	0.25	0.15	4.53	1.58	0.33	0.46	6.06	0.65	14.23	0.49	0.23
12	41.66	0.41	0.39	0.25	0.34	0.17	0.38	0.43	4.28	0.28	0.13	6.42	1.87	0.46	0.53	8.17	0.58	22.59	0.67	0.33
24	61.11	0.49	0.53	0.26	0.47	0.29	0.40	0.45	9.55	0.51	0.20	14.57	1.89	0.55	0.67	17.83	0.71	43.14	0.90	0.43

Prostate

Time (h)	R14-2	R14-4	R14-5	R14-6	R14-7	R14-8	R14-9	R14-11	R14-14	R14-15	R14-16	R14-17	R14-25	R14-29	R14-30	R14-32	R14-37	R14-43	R14-306	R14-626
1	15.09	0.33	0.38	0.26	0.40	0.14	0.28	0.26	1.28	0.26	0.13	1.56	0.37	0.38	0.47	1.48	0.54	3.08	0.65	0.21
2	27.77	0.37	0.39	0.24	0.67	0.25	0.34	0.42	1.54	0.38	0.13	2.56	0.46	0.46	0.53	2.99	0.55	6.04	0.57	0.27
4	37.41	0.39	0.39	0.21	0.44	0.30	0.40	0.51	2.61	0.52	0.16	4.69	0.53	0.41	0.58	5.05	0.48	11.89	0.56	0.40
8	55.36	0.40	0.40	0.17	0.44	0.37	0.44	0.64	4.45	0.78	0.25	8.92	0.67	0.48	0.66	9.22	0.48	23.16	0.59	0.42
12	64.17	0.74	0.63	0.30	0.42	0.56	0.49	0.59	7.11	1.03	0.39	11.97	0.69	0.61	0.73	13.48	0.54	33.97	0.60	0.43
24	77.02	1.22	0.77	0.43	0.59	0.72	0.67	0.78	16.36	1.58	0.76	32.42	0.90	0.90	0.89	27.31	0.71	58.32	0.82	0.64

Healthy

Time (h)	R14-2	R14-4	R14-5	R14-6	R14-7	R14-8	R14-9	R14-11	R14-14	R14-15	R14-16	R14-17	R14-25	R14-29	R14-30	R14-32	R14-37	R14-43	R14-306	R14-626
1	4.32	0.52	0.89	0.39	0.50	0.54	0.44	0.37	1.23	0.60	0.83	0.60	0.54	0.69	0.47	0.98	0.51	1.29	0.40	0.70
2	8.02	0.53	1.23	0.60	0.81	0.86	0.67	0.53	1.47	0.99	0.59	0.89	0.63	0.86	0.46	1.21	0.63	2.34	0.51	1.03
4	13.12	0.70	1.62	0.77	0.75	1.30	0.71	0.57	1.96	1.25	0.79	1.20	0.85	1.15	0.74	1.78	0.63	4.56	0.76	0.97
8	25.54	1.02	1.84	0.83	1.19	1.04	1.01	1.02	3.04	2.33	1.12	2.35	1.41	1.49	1.12	2.86	1.39	7.78	1.48	1.54
12	34.37	3.74	3.46	1.23	1.49	2.26	1.86	1.15	3.76	3.13	1.35	3.07	1.60	2.22	1.34	4.45	1.71	11.38	1.45	2.16
24	45.07	12.68	4.60	1.69	3.09	4.02	2.74	2.21	8.35	6.87	2.17	6.38	3.66	4.33	2.21	7.28	2.77	19.07	3.52	3.54

Figure 15. Depiction of Round 14 candidate sequences' cleavage activity in 5 separate serum pools, measured in percent cleavage.

At first glance, the top cleaving DNAzymes across all OTB cancer pools are the same as the top candidates emerging from the sensitivity analysis, namely 14-2, 14-14, 14-17, 14-32, 14-43. In fact, their cleavage values either match or exceed their values in the sensitivity analysis. 14-14 appears to conversely show higher selectivity for prostate cancer, as its degree of activity in pooled

prostate cancer serum is nearly double its activity in pancreatic, breast, colorectal, or lung cancer. The same trend is displayed by 14-17. Interestingly, with the exception of 14-14, all candidates seem to show preferential cleavage in each control cancer type compared to healthy serum, suggesting they may hold potential for general cancer detection. Further testing with additional cancer and healthy serum samples would be beneficial in exploring this observation and opportunity further.

A compelling case for pancreatic cancer specificity can be made by comparing candidate sequences' performance in fresh collected healthy participant serum to fresh collected pancreatic cancer serum. Since these serum samples have been processed identically, we are able to provide a direct cleavage comparison, without the confounding effect of serum preparation from plasma which requires additives. As such, this comparison can allow us to distinguish which sequences can detect healthy serum from disease serum, a first step in the identification of a DNAzyme sequence specific for pancreatic cancer. Unfortunately, none of the candidates identified by the sensitivity analysis show higher cleavage in pancreatic cancer serum compared to healthy serum. The highest cleavage activity in fresh collected healthy participant serum is held by 14-2 and 14-43. In fact, their cleavage values in healthy serum are nearly double their cleavage values in pancreatic cancer serum, indicating that neither of these two sequences can be used to distinguish pancreatic cancer from healthy serum. Conversely, two new sequences appear to show preferential cleavage in pancreatic cancer serum compared to healthy serum, namely **14-4** (peak cleavage 37.28% in pancreatic vs. 12.68% in healthy) and **14-15** (peak cleavage 16.61% in pancreatic vs. 6.87% in healthy). Despite their poor performance in OTB samples, these two sequences may warrant additional testing to evaluate their potential for pancreatic cancer diagnosis against healthy patients.

A possible explanation for the lower cleavage values of the candidate sequences in fresh collected pancreatic cancer serum can be provided by the observation of severe degradation in pancreatic cancer serum. Consequently, the candidate probes' performances can be flipped; rather than assuming the catalytic role and attributing its cleavage to its catalytic activity, perhaps we should be looking at the DNAzymes as substrates for endonucleases. Consequently, their performance could indicate that fresh collected pancreatic cancer serum has upregulated levels of nucleases/higher nuclease activity than normal healthy serum. This hypothesis is supported by recent advances in cancer metabolomic research which point to nucleases as molecular targets for cancer diagnosis.[12] As such, we can think of using the extent of DNAzyme digestion/degradation to tell apart pancreatic cancer serum from healthy serum, with higher values of degradation rather than cleavage pointing to pancreatic cancer.

In conclusion, this specificity analysis allowed for the emergence of two new sequences with potential for distinguishing pancreatic cancer from healthy serum: 14-4 and 14-15. Unfortunately, this analysis also demonstrated that none of the DNAzyme candidates identified in the sensitivity analysis show specificity for pancreatic cancer against other cancer types, and therefore cannot move forward into diagnostic assay assembly. However, they may hold potential for general/nonspecific cancer detection. In addition to the endonuclease effect already discussed, we propose additional factors responsible for the poor performances of the DNAzyme candidates in distinguishing pancreatic cancer from other cancers.

First, we expect considerable overlap of serum components across multiple cancer types. As such, we are likely selecting for a common biomarker. Further rounds of selection with additional selection pressures may further drive specificity towards pancreatic cancer, since the differences

between the serums explored likely come down to minute and possibly fleeting concentrations of biomarkers. This leads to the second factor, namely the possibility that rather than detecting the presence or absence of a biomarker unique to pancreatic cancer, we are detecting varying but specific concentrations of a biomarker common to all cancer types explored. This could account for the ubiquitous cleavage in all serums, but to variable degrees., and reflects a more likely reality given the complexity of serum as a reaction matrix (in addition to the complexity of the cancer metabolomic landscape). The second factor can also be applied to the nuclease hypothesis. Much like the possibility of detecting a common biomarker of varying concentrations across all cancer types, our DNAzymes may be responding to varying nuclease concentrations specific to each type of cancer. Following up with a nuclease quantifying experiment – whether through advanced size exclusion or gas chromatography-mass spectrometry, would help bring certainty to this hypothesis.

2.2.4 Kinetics of candidate RFA sequences

A kinetic analysis was performed to further characterize and evaluate the performance of each DNAzyme sequence in each serum pool. The data was gathered from the cleavage time courses discussed in the sensitivity and specificity analyses. In accordance with published work examining DNAzyme kinetics (DOI: 10.1002/anie.202012444), each reaction was modelled by non-linear regression using the one-phase association equation $Y = Ymax\ [1\text{-}e^{-kt}]$, employing the GraphPad Prism 10.0.3 software. The constraints imposed were $Y_0 = 0$, and Plateau < 100. The emerging rate constants k_{obs} have been compiled in Table 4. Their corresponding confidence intervals are additionally provided in the supplemental section.

Seq	Pancreatic Cancer	Counter Cancer Pool	Breast Cancer	Colorectal Cancer	Lung Cancer	Prostate Cancer	FC Defib Plasma – PC	FC Serum – PC	FC Serum – HP	Buffer
14-2	0.04979	0.08295	0.08048	0.07697	0.08002	0.1797	0.189	0.1999	0.08173	1.185
14-4	0.3256	1.184	1.143	0.3284	0.8586	0.06075	0.1411	0.02005	0.004547	Unstable
14-5	0.21	0.1873	0.2783	1.838	0.3355	0.2193	0.1058	0.123	0.07719	0.8596
14-6	0.0002832	0.0002503	0.00024	8.64E-05	0.000143	0.00021	0.0001732	0.001465	0.0008252	0.0002243
14-7	0.0836	0.92	0.1736	0.3948	Unstable	1.363	0.2157	0.002373	0.03261	1.395
14-8	1.792	0.3748	0.2353	0.02603	0.01167	0.1168	0.09619	0.002917	0.02608	Unstable
14-9	1.733	Unstable	2.358	0.6758	0.832	0.2885	0.4028	0.06465	0.05044	1.152
14-11	1.551	0.5042	0.8927	0.7136	0.7184	0.4309	0.8738	0.04479	0.04561	0.9619
14-14	0.004144	0.004957	0.00383	0.003953	0.004053	0.007006	0.01066	0.04757	0.01264	0.4652
14-15	1.372	0.1971	0.3512	0.3015	0.159	0.07561	0.1028	0.007178	0.002924	0.4698
14-16	0.3168	Unstable	0.4853	5.13E-05	0.07914	0.01416	0.17	0.002802	0.1022	Unstable
14-17	0.006748	0.008854	0.00634	0.007374	0.006295	0.01433	0.02554	0.002332	0.008965	4.114
14-25	0.5395	1.466	1.363	0.5837	2	0.4393	0.5991	0.004195	0.01351	3.767
14-29	0.6702	0.6947	0.6587	1.376	2.202	0.5256	0.2494	0.0358	0.02366	0.5955
14-30	3.965	1.132	1.357	1.089	0.5153	0.7467	0.3312	0.01717	0.07347	1.171
14-32	0.01008	0.006207	0.00754	0.007337	0.00795	0.0129	0.01303	0.06278	0.03795	Unstable
14-37	0.07603	0.1956	Unstable	0.4096	1.734	3.816	0.1057	0.05042	0.05903	0.03054
14-43	0.01569	0.02335	0.02294	0.0228	0.02216	0.03497	0.05071	0.0938	0.03268	0.9545
14-306	0.07856	0.4645	0.5133	0.2526	0.8868	Unstable	0.06103	0.02616	0.01068	4.103
14-626	0.07681	0.6396	Unstable	0.09958	0.6345	0.3668	0.2215	0.002476	0.05996	1.616

Table 4. Rate constants k_{obs} (hr^{-1}) of each DNAzyme candidate from Round 14 in separate serum pools. FC Defib Plasma = Fresh Collected Defibrinated Plasma. PC = Pancreatic Cancer. HP = Healthy Participant.

It can be readily observed that the buffer reactions hold the highest rate constants. While that may seem concerning at first glance, this observation must be taken in context with the plateau values defined by the model equation and constraints. In this model, we allowed the program to assign the plateau value based on the cleavage percentages obtained in the cleavage time course reaction. Consequently, bearing in mind that the rate constant a measure of how quickly a reaction reaches plateau, it is unsurprising that the buffer (control) reactions reach plateau fastest since their plateau

value is practically equivalent to the starting cleavage value. This indicates and confirms that hardly any cleavage/degradation is occurring in the buffer reactions, as evidenced by the cleavage results of the time courses as well. Ultimately, in the context of a variable plateau model, the high k_{obs} values of the DNAzymes' activities in buffer are really a reflection of their high degree of stability in buffer-only reactions at room temperature.

While there is an explanation for the higher rate constant values in buffer, this variable cleavage plateau model does pose difficulty in comparing the rate constants of different DNAzymes in each serum, as well as each DNAzyme's performance in different serum samples. For instance, it appears 14-30 has the highest rate constant in pancreatic cancer serum, suggesting it may be a viable candidate for further characterization. However, upon correlating the rate constant value with the raw cleavage data in pancreatic cancer serum, the assigned plateau of 14-30 by the equation model is only 0.424%, its 24h cleavage value is 0.47%, and the starting 1h cleavage value is 0.41%. Consequently, the rate constant will expectedly appear high since 14-30 seems to be near plateau at the first collected time point.

The difficulty in using this model points to two possibilities. First, this one-phase association model (including equation, constraints) assumes the DNAzyme cleavage is entirely target-mediated, and does not account for background degradation. Consequently, this model may not be the best suited for generating kinetic data of our DNAzyme candidates, given the complexity of serum and the high degree of non-specific interactions/degradation that is concurrently occurring. The second possibility could show us not adequately using the model to its full potential by virtue of lacking data, specifically by not extending the time course interval far enough to give more

insight into the plateau. This theory is especially poignant in fresh collected serum reactions, where the plateau does not remain stable, but rather reaches a peak cleavage amount, and the overall cleavage data of the reaction resembles a normal distribution secondary to intense degradation of the DNAzymes over time by nuclease activity. Therefore, while it may be tempting to assume that longer time courses would offer more accurate rate constant values, we must consider the significant degradation in the extended cleavage time course experiment with 14-2, in which significant degradation regardless of serum age or preparation method was observed after 24 hours.

As such, another option for addressing the difficulty in interpreting the kinetic data generated by the one-phase association model is to constrain an additional variable and attempt to normalize the k_{obs} values, in order to more readily compare them. Presently, the rate constants reflect individual/variable cleavage percentage plateaus. While the rate constant of one sequence may be higher indicating it reaches its plateau faster, the other sequence may have an overall higher cleavage plateau value and conversely a slower rate constant. This ultimately begs the question which measure is best to compare the activity of DNAzymes in variable serums. We suggest normalizing the plateau to 100% cleavage. While this is an artificial value, it does normalize the rate constant values relative to optimal DNAzyme performance (defined as 100% cleavage), allowing for easier comparison. Table 5 shows the updated k_{obs} values based on the constraint Plateau = 100.

Seq	Pancreatic Cancer	Counter Cancer Pool	Breast Cancer	Colorectal Cancer	Lung Cancer	Prostate Cancer	FC Defib Plasma – PC	FC Serum – PC	FC Serum – HP	Buffer
14-2	0.03575	0.05102	0.04801	0.05197	0.04493	0.09514	0.09567	0.02142	0.03	0.0001918
14-4	0.0002026	0.0002509	0.0002358	0.0002442	0.0002716	0.000544	0.0003222	0.02007	0.00455	0.0001864
14-5	0.0002561	0.0003035	0.0002892	0.000265	0.0002714	0.0003974	0.0003504	0.003204	0.002225	0.0002369
14-6	0.0002832	0.0002503	0.0002404	8.639E-05	0.0001434	0.0002101	0.0001732	0.001465	0.0008252	0.0002243
14-7	0.0002097	0.0002878	0.0003611	0.0001834	0.0002516	0.0003263	0.0003799	0.002373	0.001313	0.0001397
14-8	0.0001682	0.0001979	0.0003377	0.0001007	0.0001227	0.0003547	0.0003092	0.002917	0.001742	0.0002001
14-9	0.0004105	0.0004534	0.0004505	0.0001465	0.0002375	0.0003448	0.0003144	0.001274	0.001252	0.0005908
14-11	0.0002899	0.0003292	0.0004117	0.0002494	0.0002546	0.0004254	0.0002608	0.001351	0.0009882	0.000109
14-14	0.004145	0.004957	0.003826	0.003953	0.004053	0.007008	0.01066	0.003772	0.003632	0.0002044
14-15	0.000424	0.0004882	0.0005229	0.0002497	0.0002383	0.0007469	0.0007552	0.007179	0.002924	0.0001712
14-16	0.0001351	0.0002132	0.0003141	5.108E-05	9.675E-05	0.0003234	0.0006124	0.001871	0.001036	0.0002118
14-17	0.006748	0.008858	0.006337	0.007374	0.006296	0.01434	0.02556	0.002332	0.002758	0.0001759
14-25	0.0005928	0.0006125	0.0005927	0.0003763	0.001162	0.0004831	0.0003778	0.004197	0.001559	0.0002658
14-29	0.0002174	0.0003622	0.000397	0.0003099	0.0003026	0.0004459	0.0005405	0.003712	0.001894	0.0002149
14-30	0.0002732	0.000361	0.0003948	0.0002249	0.000357	0.0004907	0.0004183	0.001655	0.001036	0.0004037
14-32	0.004283	0.006207	0.007538	0.007346	0.00795	0.0129	0.01303	0.003962	0.003367	0.0002833
14-37	0.0001953	0.0002569	0.0002874	0.0002675	0.0004134	0.000386	0.0001946	0.001588	0.00129	0.00004412
14-43	0.01582	0.02335	0.02297	0.02281	0.02217	0.03498	0.04668	0.005631	0.009304	0.0001223
14-306	0.000289	0.0003642	0.0003691	0.0003685	0.0004617	0.0004452	0.0002776	0.001097	0.00149	0.00006963
14-626	0.0002066	0.000266	0.000241	0.0001589	0.0002234	0.0003272	0.0004689	0.002477	0.001642	0.00009182

Table 5. Rate constants k_{obs} (hr^{-1}) of each DNAzyme candidate from Round 14 in separate serum pools with updated kinetic model (plateau = 100). FC Defib Plasma = Fresh Collected Defibrinated Plasma. PC = Pancreatic Cancer. HP = Healthy Participant.

This new constraint added to the one-phase association model leads to the newly generated rate constants more closely resembling the results of the sensitivity and specificity analysis, whereby 14-2 appears to show the highest activity in serum and consequently holds the highest rate constant values, followed by 14-43, 14-17, 14-32, and 14-14. Predictably, it does not account for the intense degradation of DNAzymes in fresh collected serum converted from whole blood, and as such cannot accurately depict the kinetics of our DNAzymes in this medium. Fresh collected serum

activity notwithstanding, this model does more accurately reflect the low activity of DNAzymes in buffer-only reactions, and can be used to conversely identify which sequences are most stable in various media.

2.2.5 Target size parameter identification

In an effort to gain more insight into the mediating factors of DNAzyme cleavage in this selection and to narrow in on possible target(s), a size exclusion experiment was conducted. This experiment was designed to offer molecular weight parameters for possible targets, and was performed with pooled pancreatic cancer and pooled counter selection cancer serums from the OTB. Both categories of serum were centrifuged through filter columns with incrementally decreasing molecular weight cut-offs. The resulting serum fractions were used to perform cleavage time point experiments. Given the robustness of 14-2 demonstrated in this chapter, we felt it would be appropriate to continue exploring it further and employ it in this experiment. The results of this size exclusion experiment are depicted in Figure 16.

Interestingly, Figure 16 shows conclusively that cleavage of 14-2 is mediated exclusively by serum components of the largest size fraction, specifically by targets in the >100kDa size range, regardless of serum type. This finding is furthermore confirmed at two separate time points, with cleavage at 3 hours being predictably higher than at 1.5h in both serum types. Additionally, 14-2 cleavage in the largest counter selection cancer serum fraction continues to be higher than in its pancreatic cancer counterpart, consistent with 14-2's performance throughout the sensitivity, specificity, and kinetic analyses. The remaining fractions are virtually consistent with the background cleavage of 14-2, as defined by the raw value of the uncleaved marker's cleavage

band. While the results of this experiment appear conclusive, a repeat analysis with 1-2 additional DNAzymes like 14-43 and 14-14 would be beneficial in giving an indication whether the DNAzymes sequences identified in this selection all cleave in the same size fraction, or whether their cleavage is conclusively mediated by different targets. Furthermore, extending this analysis in the fresh collected pancreatic cancer samples as well as in the healthy participant samples would offer more insight into the processes responsible for DNAzyme cleavage.

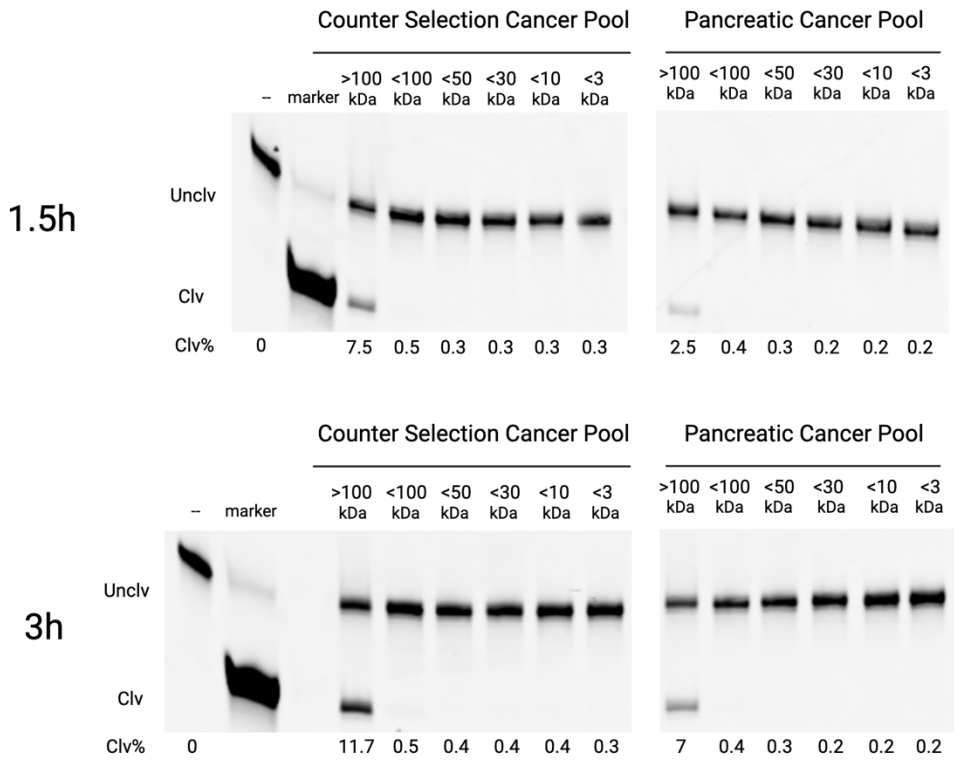


Figure 16. Gel depiction of the first two time points of the fractionated serum cleavage time point experiment.

In light of the results of the size exclusion experiment, we are left with a wide array of possible targets responsible for mediating the DNAzymes' cleavage. The finding of 14-2 cleaving in the same fraction in both counter selection cancer serum and pancreatic cancer serum gives more credence to the possibility of the DNAzymes responding to a common target found in different concentrations in pancreatic cancer compared to other types of cancer. This target may be a biomarker or more likely a serum nuclease which turns our DNAzymes into its substrates rather than allowing them to function as diagnostic signalling elements. Performing a nuclease inhibition experiment may help determine whether the DNAzyme cleavage is mediated by non-specific nuclease degradation. However, the very component responsible for inhibiting nucleases (chelating agent) will likely impair the activity of the DNAzymes as well, since both require divalent metal ions for function. Consequently, gas chromatography-mass spectrometry may prove most helpful in shedding more light on the likely target(s) responsible for DNAzyme cleavage in serum.

While it is impossible to infer the identity of the targets the DNAzymes may be responding to, we can offer suggestions based on the current literature on cancer and specifically pancreatic cancer metabolomics, as well as human serum nucleases. Table 6 shows a breakdown of pancreatic cancer-associated serum biomarkers, human serum nucleases, and their molecular weights. Those greater than 100kDa are highlighted.

Biomarker	Molecular Weight	Reference
Carbohydrate 19-9 (CA 19-9) (P)	819 Da	[13]
Carcinoembryonic Antigen (CEA)	~150-180 kDa	[14]
Hepatocyte Growth Factor (HGF) (P)	84 kDa	[15]
Exosomes	> 100 kDa	[16]
MIC-1	28 kDa	[17]
Alcohol Dehydrogenase (ADH)	81 kDa	[18]
Lactate Dehydrogenase (LDH) (monomer, dimer, tetramer)	35, 70, 140 kDa	[19]
Osteonectin	43 kDa	[20]
Osteopontin	44 kDa	[21]
IGFBP2	150 kDa	[22]
IGFBP3	28.7 – 53 kDa*	[23]
Prostate stem cell antigen	10-24 kDa	[24]
Mucin-1 (MUC1)	122 – 500 kDa*	[25]
Mucin-4 (MUC4)	550 – 930 kDa*	[26]
Mucin-13 (MUC13)	54.7 – 175 kDa*	[27]
L1CAM	200 – 220 kDa	[28]
Cancer Antigen 125 (CA-125)	110 – 2000 kDa	[29]
CEMIP	153 kDa	[30]
C4BPA (P)	67 kDa	[31]
IL-1Beta	17.5 kDa	[32]
IL-8	8.4 kDa	[33]
IL-10	37 kDa	[34]
VEGF	21 kDa	[35]
Prostate-specific antigen (PSA)	30-34 kDa	[36]
Cyclophilin b	24 kDa	[37]
Glypican-1	62 kDa	[38]
Amylase (P)	53.7 kDa	[39]
Lipase (P)	50 kDa	[40]
EGFR	180 kDa	[41]
Cancer Antigen 72-4 (CA 72-4)	220-400kDa	[42]
Tissue Polypeptide Specific antigen (TPS) (P)	20-45 kDa	[43]
Tumor-associated trypsin inhibitor (TATI) (P)	6 kDa	[44]
DUPAN-2 (P)	1000 kDa	[45]
Amylin (P)	3.9 kDa	[46]
Serum Nuclease	Molecular Weight	
DNase1	39 kDa	[47]
DNase1L3	36 kDa	[48]
Human endonuclease V (hEndoV)	50.2 kDa	[49]
Human AP-endonuclease 1 (APE1)	35 kDa	[50]
Hyaluronidase	61 kDa	[51]
RNase A	13.7 kDa	[52]
RNase H	33 kDa	[53]
RNase L	83.5 kDa	[54]
RNase T1	11 kDa	[55]
RNase 3	160 kDa	[56]
RNase 4	10-15 kDa	[57]
RNase 5 (Angiogenin)	14 kDa	[58]
RNase 7	14.5 kDa	[59]

Table 6. Biomarkers associated with cancer, followed by human serum nucleases. (P) denotes markers particularly correlated with pancreatic cancer. *size varies with degree of glycosylation.

Table 6 is not an exhaustive list of possible target candidates, yet it does offer some insights into the serum components likely responsible for mediating DNAzyme cleavage. Several biomarkers are highlighted – largely common to all cancers. Among the serum nucleases, RNase 3 appears to be most likely involved in the cleavage of the DNAzyme, presumably at the ribonucleotide unit embedded in the DNAzyme sequence. It is unclear whether upon initial cleavage by RNase 3, further digestion/degradation is taken over by the two serum DNases listed in Table 6; however, given the degree of degradation and the multitude of fluorescent bands below the DNAzyme cleavage band, this theory may have merit.

2.3 Summary

In conclusion, we have successfully demonstrated the feasibility of performing *in vitro* selection using DNAzymes directly in whole human serum. In doing so, we have shown that the *in vitro* selection process is not incapacitated by the complex metabolomics of human serum, such that a DNA library can be successfully enriched towards the identification of a target. Despite the inability to identify a DNAzyme probe highly specific for pancreatic cancer compared to other cancer types, we have identified several DNAzyme candidates showing high sensitivity for pancreatic cancer, as well as two DNAzyme candidates (14-4, 14-15) demonstrating promising specificity values compared to healthy serum. Due to the limitation posed by access to patient samples, accurate sensitivity and specificity values require further analysis. However, pending further optimization, the numerous DNAzymes identified in this study hold great potential for pancreatic cancer diagnosis and eventual incorporation into a biosensing platform, illustrating the untapped potential of *in vitro* selection in whole human serum.

2.4 Experiments

2.4.1 Enzymes, chemicals, and other reagents

Urea (ultrapure) and 40% polyacrylamide solution (29:1) were acquired from BioShop Canada (Burlington, ON, Canada). The water used was purified via Milli-Q Synthesis A10 water purifier. The enzymes T4 polynucleotide kinase (PNK), T4 DNA ligase, and Taq DNA polymerase were purchased from Thermo Scientific (Ottawa, ON, Canada). ATP, EvaGreen (20x) and deoxynucleoside 5′-triphosphates (dNTPs) were also purchased from Thermo Scientific (Ottawa, ON, Canada). Silica beads were purchased from Sigma. All other reagents were purchased from Bioshop Canada and used without further purification.

2.4.2 Synthesis and purification of oligonucleotides

A list of the oligonucleotides sequences employed in this selection experiment are listed in Table X. The DNA library PanC2, the signalling molecule FQ30, the forward PCR primer PanC2-F, the two reverse PCR primers PanC2-R and PanC2-R2, the template PanC2-Splint for ligating FQ30 to PanC2 were purchased as synthetic oligonucleotides from Integrated DNA Technologies (IDT). All oligonucleotides were purified by 10% denaturing polyacrylamide gel electrophoresis (dPAGE) prior to use and quantified spectroscopically. Each nucleotide position in the 40 nucleotide random domain of PanC2 is randomized by IDT with a 25% probability of A, C, G or T nucleotide. FQ30 contains an adenosine ribonucleotide (rA), flanked by a fluorescein-dT fluorophore and a dabcyl-dT quencher. The reverse primer PanC2-R2 contains an 18-atom spacer and a poly-T tail composed of 20 thymine nucleotides at the 5' end. The function of the spacer is to prevent the poly-T tail from being amplified, consequently marking the anti-sense strand with a lengthy poly-T tail to facilitate recovery of the DNAzyme-coding sense strand. The recovery is

accomplished through separation of the two strands by 10% dPAGE. The RNA-containing substrate FQ30 was deprotected and purified by 10% dPAGE following a previously reported protocol.[60]

Selection ID	PanC2
Library "PanC2-Lib"	5`→3` TTACGTCAAGGTGTCACTCCNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNTCATTGTTTCAGTGGCGA GCA 81nt, 40nt random domain
Forward Primer "PanC2-F"	5`→3` TTACGTCAAGGTGTCACTCC 20nt, Tm~55°C
Reverse Primer "PanC2-R"	5`→3` TGCTCGCCACTGAAACAATGA 21nt, Tm~55°C
Blocked Reverse Primer "PanC2-R2"	5`→3` TTTTTTTTTTTTTTTTTTT/iSP18/TGCTCGCCACTGAAACAAT G 41nt, Tm~55°C, iSP18 = 18 atom spacer
"FQ30" Substrate	5`→3` CTATGAACTGACXrAYGACCTCACTACCAAG 31nt, X = Dabcyl dT, rA = riboA, Y = Fluorescein dT
"PanC2-Splint" Ligation splint	5`→3` GACACCTTGACGTAACTTGGTAGTGAGGTC 30nt
i5.int.PanC2-F	5`→3` CTTTCCCTACACGACGCTCTTCCGATCTACTGTTACGTCAAGG TGTCACTCC 52nt
i7.int.PanC2-R	5`→3` GGAGTTCAGACGTGTGCTCTTCCGATCTTGCTCGCCACTGAA ACAATGA 49nt

Table 7. Oligonucleotide sequences employed in this experiment

2.4.3 Plasma sample acquisition and processing

All fifty banked plasma samples were purchased from the Ontario Tumour Board (OTB). Each sample is derived from a single patient. In total, ten Pancreatic Cancer, ten Breast Cancer, ten Colorectal Cancer, ten Lung Cancer, and ten Prostate Cancer plasma samples were obtained. Each sample amounted to 1ml of plasma. Upon receipt, samples were seroconverted in 200µl batches, with addition of 2.5µl 1M Calcium Chloride and incubation at room temperature. Samples were checked at 15 minute intervals for clotting. If no clotting was observed, an additional dose of 2.5µl 1M Calcium Chloride was added. Where clotting was not observed at 1h, 10 silica beads were added. Clotting was successfully achieved in all fifty samples in 90 minutes or less. Following clotting, samples were centrifuged at 4°C, 15000 g for 5 minutes. The supernatant (serum) was extracted and aliquoted into microcentrifuge tubes, labelled, and stored at -80°C. Prior to storage, positive and counter selection serum pools were formed and aliquoted. Positive selection serum was formed by mixing equal volumes of each of the ten pancreatic cancer patient samples. Counter selection serum was formed by mixing equal parts of each of the breast cancer, colorectal cancer, lung cancer, and prostate cancer patient samples. The pools were aliquoted, labelled, and stored at -80°C.

2.4.4. Fresh collected sample acquisition and processing

Ten fresh collected samples were acquired through recruitment of pancreatic cancer patients (5) and healthy participants (5). All participants recruited signed consent forms in accordance with the Hamilton Integrated Research Ethics Board (HIREB). Blood samples were subsequently acquired by a registered nurse or the writer, using standard blood collection vacutainer tubes and 21G butterfly needles. Approximately 10ml of blood was collected in total, into two vacutainer tubes.

Nine samples were collected in red-capped serum collection tubes, and one sample was acquired in a purple-capped EDTA coated collection tube. All five healthy participant samples were collected in red-capped serum tubes. Upon collection, the samples were couriered to the laboratory facility. The pancreatic cancer sample collected in the EDTA-coated tube was processed to plasma by centrifuging the vacutainer tubes in a swing-bucket centrifuge at 4°C, 15000 g for 15 minutes to pellet cells. The plasma supernatant was transferred to cryovials. The vials were labelled with a non-patient identifying ID, date of collection, and stored at -80°C. Seroconversion followed the banked sample protocol detailed in section 2.5.3. All remaining blood samples collected in serum vacutainers were allowed to clot at room temperature for 30 minutes. The vacutainers were centrifuged at 4°C, 2200 g for 20 minutes. The supernatant (serum) was aliquoted into microcentrifuge tubes, labelled, and stored at -80°C.

2.4.5. in vitro selection

The *in vitro* selection protocol followed previously described protocols by our research group. Briefly, 1300pmol of PanC2 was phosphorylated with ATP and 20U of T4 polynucleotide kinase (PNK) in 1× PNK buffer A at 37°C for 20 minutes, in a 100µl reaction volume. The reaction was stopped by heating the mixture at 90°C for 5 minutes. Water, along with equimolar FQ30 and PanC2-Splint were subsequently added to this solution, to a volume of 366µl, followed by heating at 90°C for 1 min and cooled to room temperature to anneal the fragments. Next, 30µl of 10× T4 DNA ligase buffer was added, followed by T4 DNA ligase (20 U) was added, to a total volume of 400µl. The reaction was subsequently incubated at room temperature for 2 h. The DNA molecules in the mixture were concentrated by ethanol precipitation, and the ligated PanC2-FQ30 constructs were purified by 10% dPAGE and quantified.

The first round of selection was a positive selection round. Consequently, all 1300pmol of the PanC2-FQ30 ligated DNAzyme library was mixed with water, heated at 90°C for 3 minutes and cooled at ambient temperature for 5 minutes to allow DNAzyme sequences to fold. 10x Selection Buffer was added to the reaction (500mM HEPES pH 7.0, 1500mM NaCl, 150mM $MgCl_2$) to a final concentration of 1x, followed by addition of the positive selection serum in accordance with the selection conditions detailed in Table 1. The reaction was stopped by addition of 2x Quenching Buffer (2x Urea Loading Buffer + 60mM EDTA), to a final concentration of 1x. The cleaved PanC2-FQ30 sequences were purified by 10% dPAGE, and resuspended in 20µl of water. Each selection round gel was scanned using the Amersham Typhoon Biomolecular Imager.

Following positive selection, two PCR reactions (PCR1 and PCR2) were performed to amplify and recover the DNAzyme-coding sense strand. PCR1 was performed with 0.5µM PanC2-F and 0.5µM PanC2-R primers and 5µl of the purified cleaved PanC2-FQ30 from the previous round of positive selection, along with 200 µM each of dNTPs (dATP, dCTP, dGTP and dTTP), 1× PCR buffer (75 mM Tris-HCl, pH 9.0, 2 mM MgCl2, 50 mM KCl, 20 mM (NH4)2SO4) and 2.5 U of Thermus aquaticus (Taq) DNA polymerase. The DNA was amplified using the following thermocycling steps: 95°C for 1 min; 14-16 (dependent on the amount of cleavage of the PanC2 DNA library) cycles of 95°C for 20s, 52°C for 30s and 72°C for 30s; 72°C for 1 min. For the PCR2 reaction, 5µl of a 1:100 dilution of the PCR1 product was used as the template for this additional PCR step (a total of 48× 50-µL reactions were conducted to generate enough DNA) using primers PanC2-F and PanC2-R2 and the same protocol as PCR1. The sense strand was purified by 10% dPAGE and used for the next selection round.

Counter selection reactions followed the same protocol as positive selection, with the uncleaved PanC2-FQ30 construct instead being purified by dPAGE and carried into the next positive selection round.

2.4.6 Cleavage Time Course Reactions

Cleavage time course reactions with Selection Buffer (SB), Counter Selection Cancer Serum Pool and Pancreatic Cancer Serum Pool comparing the starting library (R0) to the recovered library from Round 5 (R5) were carried out as follows: 30pmol PanC2-FQ30 constructs from R0 and R5 were incubated with 30%v/v serum, 6µl of 10× SB, and water to a reaction volume of 60 µl. The buffer only time course substituted serum with an equal volume of water. Time courses were initiated by addition of all components, and incubated at ambient temperature draped with a paper towel for 48 hours. At the indicated time points, a 10µl aliquot was removed from each reaction and mixed with 10µl 2x Quenching Buffer (QB). Time point aliquots were stored at -20°C until completion of the time course. The cleavage was then analyzed by 10% dPAGE. The image of cleaved and uncleaved DNAzyme bands was obtained with the Amersham Typhoon Biomolecular Imager.

2.4.7 High-throughput sequencing

The cleavage product from the rounds 6-14 was amplified by PCR to obtain sufficient DNA for sequencing. PCR1 was conducted using PanC2-F and PanC2-R following the same protocol as described above. 5µL of the 1:100 diluted PCR1 product was was used as the template for PCR2 using deep sequencing internal primers i5.int.PanC2-F/i7.int.PanC2-R, and again with assigned external sequencing primers, using the same protocol above for PCR1. 4 individual external primer

PCR reactions for each recovered library were performed, and the PCR products were purified by 2% agarose gel electrophoresis. DNA extraction from agarose gel was done using Monarch® DNA Gel Extraction Kit (New England BioLabs). Purified PCR products were sequenced using paired-end Next-Generation sequencing (NGS) using an Illumina Miseq system at the Farncombe Metagenomics Facility, McMaster University. Raw sequencing reads were first trimmed of their primers using Geneious. The resulting 40 nt reads were filtered for quality using PrinSeq v0.20.4 to make sure only high-quality reads were used for further analysis. 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. 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.

2.4.8 Sensitivity and Specificity Analysis

Following identification of candidate sequences based on enrichment trends identified in the sequencing data, the candidate DNAzyme sequences were ordered from IDT, ligated with the FQ30 segment, and employed in cleavage time course reactions with individual and pooled patient samples. Both the banked samples from the OTB and the fresh collected samples were tested in the sensitivity and sequencing analysis. The protocol followed the cleavage time course protocol previously described, with time point collection extended to 1h – 96h.

2.4.9 Kinetics Analysis

A kinetic analysis was performed to identify the rate constants of candidate DNAzymes in various patient samples. The data was gathered from the cleavage time courses discussed in the sensitivity and specificity analyses. In accordance with published work examining DNAzyme kinetics (DOI: 10.1002/anie.202012444), each reaction was modelled by non-linear regression using the one-phase association equation $Y = Ymax [1-e^{-kt}]$, employing the GraphPad Prism 10.0.3 software. The constraints imposed were $Y_0 = 0$, and Plateau < 100. A second analysis was performed with the same model and constraints, only with Plateau = 100.

2.4.10 Size Exclusion Experiment

A size exclusion experiment was performed using the most active candidate from the sensitivity and specificity testing (14-2). Four categories of serum were tested, and include: Pooled OTB Pancreatic cancer samples, Pooled OTB Counter Selection Cancer samples, Pooled Fresh Collected Healthy Participant Samples, and Pooled Fresh Collected Pancreatic Cancer samples. Each pooled serum category was fractionated using size exclusion columns with the following filter size parameters: 100kDa, 50kDa, 30kDa, 10kDa, 3kDa. Beginning with 200µl of each serum pool, serial fractioning was completed by centrifuging the serum with each size column at 4°C, 14000g, for 5 minutes. The concentrated fraction above the filter was recovered, and the eluent was transferred to the next column in decreasing order of filter size. Cleavage time course reactions were performed by incubating the ligated 14-2 DNAzymes with each recovered serum fraction. The cleavage was then analyzed by 10% dPAGE. The image of cleaved and uncleaved DNAzyme bands was obtained with the Amersham Typhoon Biomolecular Imager.

2.5 References

1. Zhang Z, Li J, Gu J, et al. A Universal DNA Aptamer that Recognizes Spike Proteins of Diverse SARS-CoV-2 Variants of Concern. *Chemistry – A European Journal*. 2022;28(15). doi:10.1002/chem.202200078
2. Li J, Khan S, Gu J, Filipe CDM, Didar TF, Li Y. A Simple Colorimetric Au-on-Au Tip Sensor with a New Functional Nucleic Acid Probe for Food-borne Pathogen *Salmonella typhimurium*. *Angewandte Chemie International Edition*. 2023;62(20). doi:10.1002/anie.202300828
3. Qian S, Chang D, Gu J, Salena BJ, Li Y. In Vitro Selection and Characterization of a DNAzyme Probe for Diverse Pathogenic Strains of *Clostridium difficile*. *Chemistry – A European Journal*. 2023;29(36). doi:10.1002/chem.202300240
4. Qian S, McConnell EM, Rothenbroker M, et al. Detecting *Legionella pneumophila* in Cooling Tower Water Samples with a DNAzyme/Bead-Based Fluorescence Assay. *Analysis & Sensing*. Published online May 25, 2023. doi:10.1002/anse.202300020
5. Feng Q, Zakaria S, Morrison D, et al. A Fluorogenic DNAzyme for A Thermally Stable Protein Biomarker from *Fusobacterium nucleatum* , a Human Bacterial Pathogen. *Angewandte Chemie International Edition*. 2023;62(35). doi:10.1002/anie.202306272
6. Gold L. SELEX: How It Happened and Where It will Go. *J Mol Evol*. 2015;81(5-6):140-143. doi:10.1007/s00239-015-9705-9
7. Shen Z, Wu Z, Chang D, et al. A Catalytic DNA Activated by a Specific Strain of Bacterial Pathogen. *Angewandte Chemie International Edition*. 2016;55(7):2431-2434. doi:10.1002/anie.201510125
8. Liu M, Chang D, Li Y. Discovery and Biosensing Applications of Diverse RNA-Cleaving DNAzymes. *Acc Chem Res*. 2017;50(9):2273-2283. doi:10.1021/acs.accounts.7b00262
9. National Institute of Health, National Cancer Institute. Cancer Stat Facts: Common Cancer Sites. Surveillance, Epidemiology, and End Results Program.
10. GREEN D. Coagulation cascade. *Hemodialysis International*. 2006;10(S2). doi:10.1111/j.1542-4758.2006.00119.x

11. Castro AR, Kikkert SE, Fears MB, Pope V. Defibrination of Blood Plasma for Use in Serological Tests for Syphilis. *Clinical and Vaccine Immunology*. 2002;9(6):1376-1378. doi:10.1128/CDLI.9.6.1376-1378.2002
12. Balian A, Hernandez FJ. Nucleases as molecular targets for cancer diagnosis. *Biomark Res*. 2021;9(1):86. doi:10.1186/s40364-021-00342-4
13. Borenstein-Katz A, Warszawski S, Amon R, et al. Biomolecular Recognition of the Glycan Neoantigen CA19-9 by Distinct Antibodies. *J Mol Biol*. 2021;433(15):167099. doi:10.1016/j.jmb.2021.167099
14. Mitchell KennethF. A carcinoembryonic antigen (CEA) specific monoclonal hybridoma antibody that reacts only with high-molecular-weight CEA. *Cancer Immunology Immunotherapy*. 1980;10(1):1-5. doi:10.1007/BF00199272
15. NAKAMURA T, MIZUNO S. The discovery of Hepatocyte Growth Factor (HGF) and its significance for cell biology, life sciences and clinical medicine. *Proceedings of the Japan Academy, Series B*. 2010;86(6):588-610. doi:10.2183/pjab.86.588
16. Qadir F, Aziz MA, Sari CP, et al. Transcriptome reprogramming by cancer exosomes: identification of novel molecular targets in matrix and immune modulation. *Mol Cancer*. 2018;17(1):97. doi:10.1186/s12943-018-0846-5
17. Bauskin AR, Zhang HP, Fairlie WD, et al. The propeptide of macrophage inhibitory cytokine (MIC-1), a TGF-β superfamily member, acts as a quality control determinant for correctly folded MIC-1. *EMBO J*. 2000;19(10):2212-2220. doi:10.1093/emboj/19.10.2212
18. Millar AA, Olive MR, Dennis ES. The expression and anaerobic induction of alcohol dehydrogenase in cotton. *Biochem Genet*. 1994;32(7-8):279-300. doi:10.1007/BF00555831
19. Jafary F, Ganjalikhany MR, Moradi A, Hemati M, Jafari S. Novel Peptide Inhibitors for Lactate Dehydrogenase A (LDHA): A Survey to Inhibit LDHA Activity via Disruption of Protein-Protein Interaction. *Sci Rep*. 2019;9(1):4686. doi:10.1038/s41598-019-38854-7
20. Guweidhi A, Kleeff J, Adwan H, et al. Osteonectin Influences Growth and Invasion of Pancreatic Cancer Cells. *Ann Surg*. 2005;242(2):224-234. doi:10.1097/01.sla.0000171866.45848.68
21. Subramani V. OPN –Revisited. *JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH*. Published online 2015. doi:10.7860/JCDR/2015/12872.6111

22. Wood WI, Cachianes G, Henzel WJ, et al. Cloning and Expression of the Growth Hormone-Dependent Insulin-Like Growth Factor-Binding Protein. *Molecular Endocrinology*. 1988;2(12):1176-1185. doi:10.1210/mend-2-12-1176
23. Varma Shrivastav S, Bhardwaj A, Pathak KA, Shrivastav A. Insulin-Like Growth Factor Binding Protein-3 (IGFBP-3): Unraveling the Role in Mediating IGF-Independent Effects Within the Cell. *Front Cell Dev Biol*. 2020;8. doi:10.3389/fcell.2020.00286
24. Reiter RE, Gu Z, Watabe T, et al. Prostate stem cell antigen: A cell surface marker overexpressed in prostate cancer. *Proceedings of the National Academy of Sciences*. 1998;95(4):1735-1740. doi:10.1073/pnas.95.4.1735
25. Lan MS, Batra SK, Qi WN, Metzgar RS, Hollingsworth MA. Cloning and sequencing of a human pancreatic tumor mucin cDNA. *J Biol Chem*. 1990;265(25):15294-15299.
26. Chaturvedi P, Singh AP, Batra SK. Structure, evolution, and biology of the MUC4 mucin. *The FASEB Journal*. 2008;22(4):966-981. doi:10.1096/fj.07-9673rev
27. Maher DM, Gupta BK, Nagata S, Jaggi M, Chauhan SC. Mucin 13: Structure, Function, and Potential Roles in Cancer Pathogenesis. *Molecular Cancer Research*. 2011;9(5):531-537. doi:10.1158/1541-7786.MCR-10-0443
28. Kiefel H, Bondong S, Hazin J, et al. L1CAM. *Cell Adh Migr*. 2012;6(4):374-384. doi:10.4161/cam.20832
29. Weiland F, Fritz K, Oehler MK, Hoffmann P. Methods for Identification of CA125 from Ovarian Cancer Ascites by High Resolution Mass Spectrometry. *Int J Mol Sci*. 2012;13(8):9942-9958. doi:10.3390/ijms13089942
30. Spataro S, Guerra C, Cavalli A, et al. <scp>CEMIP</scp> (<scp>HYBID</scp> , <scp>KIAA1199</scp>): structure, function and expression in health and disease. *FEBS J*. 2023;290(16):3946-3962. doi:10.1111/febs.16600
31. Sogawa K, Yamanaka S, Takano S, et al. Fucosylated C4b-binding protein α-chain, a novel serum biomarker that predicts lymph node metastasis in pancreatic ductal adenocarcinoma. *Oncol Lett*. 2020;21(2):127. doi:10.3892/ol.2020.12388
32. Yaseen MM, Abuharfeil NM, Darmani H. The role of IL-1β during human immunodeficiency virus type 1 infection. *Rev Med Virol*. 2023;33(1). doi:10.1002/rmv.2400

33. Qazi BS, Tang K, Qazi A. Recent Advances in Underlying Pathologies Provide Insight into Interleukin-8 Expression-Mediated Inflammation and Angiogenesis. *Int J Inflam*. 2011;2011:1-13. doi:10.4061/2011/908468
34. Minshawi F, Lanvermann S, McKenzie E, et al. The Generation of an Engineered Interleukin-10 Protein With Improved Stability and Biological Function. *Front Immunol*. 2020;11. doi:10.3389/fimmu.2020.01794
35. Leung DW, Cachianes G, Kuang WJ, Goeddel D V., Ferrara N. Vascular Endothelial Growth Factor Is a Secreted Angiogenic Mitogen. *Science (1979)*. 1989;246(4935):1306-1309. doi:10.1126/science.2479986
36. Tang Y, Cui Y, Zhang S, Zhang L. The sensitivity and specificity of serum glycan-based biomarkers for cancer detection. In: ; 2019:121-140. doi:10.1016/bs.pmbts.2019.01.010
37. DeBoer J, Madson CJ, Belshan M. Cyclophilin B enhances HIV-1 infection. *Virology*. 2016;489:282-291. doi:10.1016/j.virol.2015.12.015
38. Belting M, Mani K, Jönsson M, et al. Glypican-1 Is a Vehicle for Polyamine Uptake in Mammalian Cells. *Journal of Biological Chemistry*. 2003;278(47):47181-47189. doi:10.1074/jbc.M308325200
39. Bano S, Qader SAU, Aman A, Syed MN, Azhar A. Purification and Characterization of Novel α-Amylase from Bacillus subtilis KIBGE HAS. *AAPS PharmSciTech*. 2011;12(1):255-261. doi:10.1208/s12249-011-9586-1
40. Febriani, Aura N, Kemala P, Saidi N, Iqbalsyah TM. Novel thermostable lipase produced by a thermo-halophilic bacterium that catalyses hydrolytic and transesterification reactions. *Heliyon*. 2020;6(7):e04520. doi:10.1016/j.heliyon.2020.e04520
41. Turk HF, Chapkin RS. Analysis of Epidermal Growth Factor Receptor Dimerization by BS3 Cross-Linking. In: ; 2015:25-34. doi:10.1007/978-1-4939-1789-1_3
42. Li M, Xue F, Yang J, Pan X. Correlation between tumor marker CA72-4 and prognosis of patients with gastric cancer. *Medicine*. 2020;99(52):e23723. doi:10.1097/MD.0000000000023723
43. Cetin T, Oğuz A, Algan P, Yildirim IS. The use of tissue polypeptide specific antigen as a marker in liver diseases. *Turk J Gastroenterol*. 2003;14(3):177-180.
44. Stenman UH. Tumor-associated trypsin inhibitor. *Clin Chem*. 2002;48(8):1206-1209.

45. Lan MS, Khorrami A, Kaufman B, Metzgar RS. Molecular characterization of a mucin-type antigen associated with human pancreatic cancer. The DU-PAN-2 antigen. *Journal of Biological Chemistry*. 1987;262(26):12863-12870. doi:10.1016/S0021-9258(18)45286-6
46. Adeghate E, Kalász H. Amylin Analogues in the Treatment of Diabetes Mellitus: Medicinal Chemistry and Structural Basis of its Function. *Open Med Chem J*. 2011;5(Suppl 2):78-81. doi:10.2174/1874104501105010078
47. Jiskoot W, Crommelin DJA. Biophysical and Biochemical Characteristics of Therapeutic Proteins. In: *Pharmaceutical Biotechnology*. Springer International Publishing; 2019:19-32. doi:10.1007/978-3-030-00710-2_2
48. Napirei M, Wulf S, Eulitz D, Mannherz HG, Kloeckl T. Comparative characterization of rat deoxyribonuclease 1 (Dnase1) and murine deoxyribonuclease 1-like 3 (Dnase1l3). *Biochemical Journal*. 2005;389(2):355-364. doi:10.1042/BJ20042124
49. Mi R, Alford-Zappala M, Kow YW, Cunningham RP, Cao W. Human endonuclease V as a repair enzyme for DNA deamination. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2012;735(1-2):12-18. doi:10.1016/j.mrfmmm.2012.05.003
50. Li M, Wilson DM. Human Apurinic/Apyrimidinic Endonuclease 1. *Antioxid Redox Signal*. 2014;20(4):678-707. doi:10.1089/ars.2013.5492
51. Lee BM, Park SJ, Noh I, Kim CH. The effects of the molecular weights of hyaluronic acid on the immune responses. *Biomater Res*. 2021;25(1):27. doi:10.1186/s40824-021-00228-4
52. Larson SB, Day JS, Nguyen C, Cudney R, McPherson A. Structure of bovine pancreatic ribonuclease complexed with uridine 5′-monophosphate at 1.60 Å resolution. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2010;66(2):113-120. doi:10.1107/S174430910905194X
53. WU H, LIMA WF, CROOKE ST. Molecular Cloning and Expression of cDNA for Human RNase H. *Antisense Nucleic Acid Drug Dev*. 1998;8(1):53-61. doi:10.1089/oli.1.1998.8.53
54. Tiev KP, Demettre E, Ercolano P, Bastide L, Lebleu B, Cabane J. RNase L Levels in Peripheral Blood Mononuclear Cells: 37-Kilodalton/83-Kilodalton Isoform Ratio Is a Potential Test for Chronic Fatigue Syndrome. *Clinical and Vaccine Immunology*. 2003;10(2):315-316. doi:10.1128/CDLI.10.2.315-316.2003

55. Dudkina E, Kayumov A, Ulyanova V, Ilinskaya O. New Insight into Secreted Ribonuclease Structure: Binase Is a Natural Dimer. *PLoS One*. 2014;9(12):e115818. doi:10.1371/journal.pone.0115818
56. Blank A, Dekker CA. Ribonucleases of human serum, urine, cerebrospinal fluid, and leukocytes. Activity staining following electrophoresis in sodium dodecyl sulfate-polyacrylamide gels. *Biochemistry*. 1981;20(8):2261-2267. doi:10.1021/bi00511a030
57. Terzyan SS, Peracaula R, de Llorens R, et al. The three-dimensional structure of human RNase 4, unliganded and complexed with d(up), reveals the basis for its uridine selectivity. *J Mol Biol*. 1999;285(1):205-214. doi:10.1006/jmbi.1998.2288
58. Fasoli S, Bettin I, Montioli R, et al. Dimerization of Human Angiogenin and of Variants Involved in Neurodegenerative Diseases. *Int J Mol Sci*. 2021;22(18):10068. doi:10.3390/ijms221810068
59. Rademacher F, Dreyer S, Kopfnagel V, Gläser R, Werfel T, Harder J. The Antimicrobial and Immunomodulatory Function of RNase 7 in Skin. *Front Immunol*. 2019;10. doi:10.3389/fimmu.2019.02553
60. Ali MM, Aguirre SD, Lazim H, Li Y. Fluorogenic DNAzyme Probes as Bacterial Indicators. *Angewandte Chemie International Edition*. 2011;50(16):3751-3754. doi:10.1002/anie.201100477

Chapter 3. Identification of RNA-cleaving Fluorogenic DNAzymes for Detection of Malignant Hyperthermia Susceptibility in Human Serum

3.1 Introduction

An important observation from the pancreatic cancer selection became apparent fairly early in its progression – namely the challenge of performing a selection on serum from an active pathology, and one as complex as cancer. Consequently, a decision was made to re-explore *in vitro* selection in whole serum and apply it to a non-acute pathology with a less heterogeneous biomarker landscape than cancer. Malignant hyperthermia (MH) was subsequently selected as the pathology of interest, also reflecting the clinical interests of the writer.

It is however worth noting that any selection in whole serum is still expected to hold a high degree of complexity in the molecular interactions of the DNAzyme library with the serum components. Additionally, while there are fewer genes involved in the pathophysiologic mechanism and development of MH compared to cancer,[1] genome sequencing has uncovered more than 200 mutations in those genes associated with malignant hyperthermia.[2] As such, in addition to individual-to-individual serum variability, the serum samples used for this selection are still expected to garner significant heterogeneity secondary to genotypic differences and variable penetrance of these mutations,[1] as reflected in the variable clinical thresholds for triggering an MH reaction.[3]

Our hope in pursuing a selection in MH sensitive serum is to further establish the feasibility of performing *in vitro* selection in whole serum, in this case applying it to a disorder with a more conventional positive/negative diagnostic mechanism rather than an acute and active disease process such as a tumor with a complex associated microenvironment, the latter of which is more challenging to control for when designing experiments. Additionally, we aim to identify

DNAzyme sequences capable of accurately detecting malignant hyperthermia susceptibility in individual serum samples, along with the DNAzymes' associated targets/biomarkers. The long-term goal of this work would once again see the development of a less-invasive FNA-based diagnostic tool for point-of-care detection of MH susceptibility.

3.1.1 Study design

Before proceeding to discuss the selection plan/study design, an important distinction must be made between malignant hyperthermia (MH) and malignant hyperthermia susceptibility (MHS). MH refers to the acute hypermetabolic reaction triggered in an individual with the genetic predisposition following exposure to an anesthetic trigger.[4] MHS however, refers to the non-acute state of the individual carrying the genetic traits that would predispose him/her/them to an MH reaction if exposed to the triggering agents by way of the dysfunctional ryanodine receptor (RYR1).[5,6]

Since the goal of the project is to develop a serum-based diagnostic assay to serve as a viable less-invasive alternative to the CHCT gold standard test[7], we reduced the diagnosis of MHS to a yes/no mechanism as indicated by the results of the CHCT. Given that the CHCT thresholds for detecting MH susceptibility align with observed clinical thresholds for triggering a reaction, by aligning our positive/negative test parameters to the CHCT parameters, we hope to construct an assay whose positive tests are indicative of clinical significance, and can therefore be used in a point-of-care manner to guide anesthetic management.

We must however acknowledge that reducing MHS diagnosis to a positive/negative dichotomy does not necessarily reflect the most accurate mechanism of MH pathophysiology. Based on the

number of genes involved, their degree of penetrance and associated mutations, the pathophysiology of MH would also likely yield variable levels of biomarkers. We do however posit that positive CHCT results would likely correlate with more variety in the biomarker landscape, or with higher concentrations of potential MHS-specific biomarkers, both of which can be detected and differentiated with DNAzymes. Therefore, we are prepared to find that once again we may be identifying DNAzymes that respond to biomarkers common to both our positive and counter selection pools, but in differing concentrations. Consequently, we understand that our protocol may lead to the detection of subclinical MH susceptibility in the control samples, and are therefore prepared to contextualize potentially low specificity values by relating them back to the clinical diagnosis, CHCT performance, and genetic information where available. Nevertheless, this experiment bears repeating given the great need for an alternative to MH susceptibility detection, and to bring further validation to the feasibility of performing *in vitro* selection in whole serum from human samples for diagnostic application.

When designing this selection plan, we employed a virtually identical protocol to the pancreatic cancer selection, as can be visualized in Figure 1. We retained the same library design, but with differences in the sequences of the constant primer regions. This was done to ensure no contamination between selections could occur and confound the results. We also maintained the same reaction components and concentrations, including the same signal generating molecule (FQ30). Following 10+ rounds of selection, the enriched library was also sequenced using the Illumina MiSeq Next Generation Sequencing method to identify potential DNAzyme candidates. These candidates subsequently underwent sensitivity, specificity, and kinetic analysis, followed by a size exclusion experiment to offer some insights on a potential target.

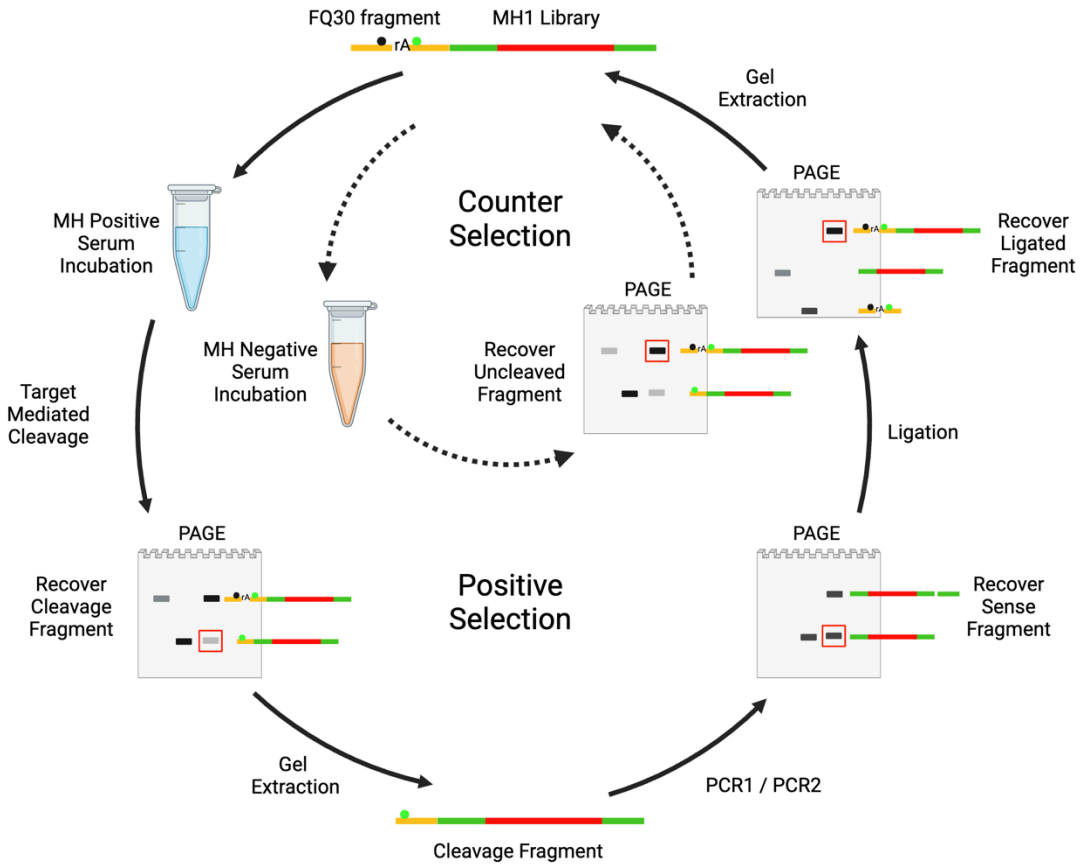


Figure 1. *in vitro* Selection/SELEX scheme with the MH1 Library using MHS (positive), and MHN (counter) serum. Counter selection and positive selection cycles are depicted, performed in iterative series.

Where the pancreatic cancer and MH selection plans diverge is in the incubation times for positive and counter selection reactions, and in the sensitivity and specificity analysis design. First, we elected to increase the positive selection incubation time to 4 hours to allow for a more robust start to the selection, and avoid paring down the DNAzyme library excessively from the onset of the protocol. While this does come at the cost of sacrificing some affinity/sensitivity at the beginning of the selection, this attribute can be optimized later in the protocol by decreasing the incubation time. We additionally opted for a much longer counter selection incubation time given the difficulties with specificity in the pancreatic cancer selection. Unlike sensitivity which can be

improved more readily, specificity is more challenging to rein in once the selection protocol is underway due to the significant amplification of sequences in each round. Consequently, we assigned a variable counter selection incubation time interval between 16-24 hours. Given that *in vitro* selection is a dynamic process, we hypothesized that a variable counter selection incubation time may prevent the selection conditions from stagnating and becoming quickly exhausted. A summary of the selection conditions for each round of selection performed are detailed in Table 1.

The second aspect of the protocol we were able to greatly improve upon was the sensitivity and specificity analysis design, by virtue of access to a higher volume of individual patient samples. We were able to obtain 45 MHS and 45 MHN samples to perform our validation studies, for a total of 90 patient samples, all of which received confirmatory testing using the gold standard CHCT at the Malignant Hyperthermia Investigation Unit (MHIU) at the University of Toronto. Performing sensitivity and specificity studies with such a large number of patient samples will naturally impact our ability to test a variety of candidate sequences. Therefore, unlike the pancreatic cancer selection where we were able to evaluate 20 candidates, in this study we employed the standard approach of testing the top 4 sequences in the final selection round. The benefit of having robust access to patient samples for validation and generalizability of DNAzyme performance cannot be understated, especially compared to the pancreatic cancer project, where validation was performed in the same samples used for the *in vitro* selection protocol and only 10 new fresh collected samples.

Round	Round Type	[Target]	Target Sample	[Library] (nM)	Library Input (pmol)	Reaction Time (hours)
1	+	30%	MHS Pool	5000	2000	4
2	-	1X	Selection Buffer	1000	200	16
2	+	30%	MHS Pool	1000	100	4
3	-	1X	Selection Buffer	500	200	24
3	+	30%	MHS Pool	500	100	4
4	-	30%	MHN Pool	500	200	16
4	+	30%	MHS Pool	500	100	4
5	-	30%	MHN Pool	500	200	16
5	+	30%	MHS Pool	500	100	4
6	-	30%	MHN Pool	500	200	24
6	+	30%	MHS Pool	500	100	4
7	-	30%	Counter MH Pool	500	200	24
7	+	30%	MHS Pool	500	100	4
8	-	30%	MHN Pool	500	200	16
8	+	30%	MHS Pool	500	100	4
9	-	30%	MHN Pool	500	200	16
9	+	30%	MHS Pool	500	100	4
10	-	30%	MHN Pool	500	200	16
10	+	30%	MHS Pool	500	100	4
11	-	30%	MHN Pool	500	200	24
11	+	30%	MHS Pool	500	100	4
12	-	30%	MHN Pool	500	200	24
12	**+**	**30%**	**MHS Pool**	**500**	**100**	**1**

Table 1. Selection conditions for all 12 rounds of positive (+) and counter (-) selection using the MH1 DNA library.

3.1.2 Clinical sample procurement and processing

We were very fortunate to be able to collaborate with the MHIU at the University of Toronto. Subsequently, all 90 MH sensitive and MH negative plasma samples were provided by the MHIU from their biobank, in three installments. The first 20 samples (10 MHS/10 MHN) sent were seroconverted and combined to create the positive and counter selection serum pools for initiation of *in vitro* selection. The additional 70 plasma samples (35 MHS/ 35 MHN) were sent at the validation stage of the study, in two installments. Every sample we received was initially obtained by MHIU staff at the time of each referred patient undergoing their MH investigation via CHCT. Consequently, every patient's diagnosis is made according to the gold standard. For each sample, we were provided with a comprehensive clinical background and indication for testing, facilitating our ability to contextualize our results, particularly during the sensitivity and specificity section. The compiled clinical data is provided in the supplemental section. Of note, attempts to seroconvert the 70 new plasma samples using the same calcium chloride protocol proved unsuccessful. In retrospect, this was likely attributed to higher amounts of EDTA present in the plasma of these samples, transferred in from the blood collection tube. Subsequently, we devised a new seroconversion protocol using 25U of Thrombin per 200 microliters of plasma to achieve clotting, followed by centrifugation and aliquoting of the emerging serum fraction. All 70 new samples were successfully seroconverted using Thrombin.

3.2 Results

3.2.1 *in vitro* selection

The MH1 library underwent ligation with the FQ30 segment, generating 800pmol of ligated product. This product was carried into the first round of positive selection, following which we entered a cycle of alternating rounds of counter and positive selection. The first two counter selection rounds were once again performed in selection buffer with no serum in order to eliminate any self-cleaving sequences, whether activated by selection buffer or not. Beginning with round 4, counter selection was carried out with MHN serum. Each selection reaction was then run on an SDS-PAGE gel, which was subsequently imaged on a fluorescence scanner to identify the cleaved and uncleaved DNAzyme bands. According to the type of selection reaction (positive/negative), the corresponding band of interest was excised and purified. The fluorescence emissions of the visualized DNAzyme gel bands in each selection reaction were quantified by a fluorescence-interpreting software, and used to calculate a percent cleavage value for every selection reaction. The emerging percent cleavage values are presented in Figure 2.

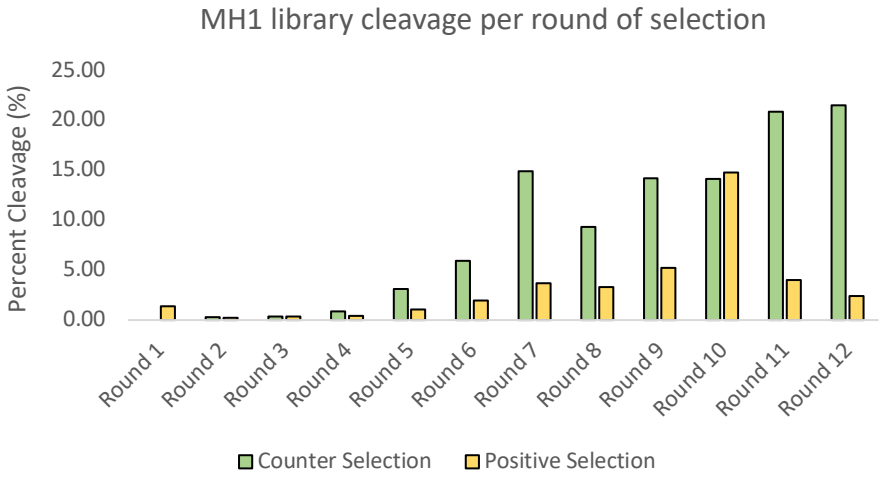


Figure 2. Graph representation of the percent cleavage of DNAzyme library for Rounds 1-12 of positive and counter selection.

As expected, the first two counter selection rounds demonstrate minimal cleavage, with an uptick at Round 4, when MHN serum was introduced into the selection. Within counter selection, the cleavage rates vary across the rounds but appear to correlate with the incubation time, which is an early indication of a time-dependent target interaction. Unsurprisingly, counter selection cleavage values exceed positive selection cleavage values with the introduction of MHN serum, given the extended counter selection incubation time. Interestingly, the positive selection cleavage appears to peak at Round 10, followed by a sharp decline at Round 11, following an extended 24 hour counter selection incubation time. This would suggest that the extended counter selection incubation may have removed more cross-reactive sequences from the DNAzyme pool, leaving fewer sequences behind to carry into positive selection. This observation points to enrichment of the library, however several other signs of enrichment are present and merit highlighting.

Beginning at Round 4, we can detect steady incremental increase in DNAzyme library cleavage in both counter and positive selection reactions, suggesting round-to-round enrichment is occurring. Furthermore, comparing the cleavage values in the counter selection reactions of Round 7 and 11 demonstrates increased cleavage in Round 11 for the same incubation time; this comparison can essentially serve as a proxy to a cleavage time course reaction. By Round 12 however, it appears that positive selection cleavage is plateauing, suggesting possible exhaustion and loss of enrichment of the library under the current selection conditions. Consequently, a decision was made to sequence the Round 12 library and the enriched libraries from the 6 rounds preceding it for comparison.

In addition to the enriched libraries from Rounds 6-12, we performed a cleavage time point experiment with the ligated product of the Round 12 enriched library (effectively Round 13), and sent each time point aliquot for sequencing. The results of the time course are depicted in Figure 3.

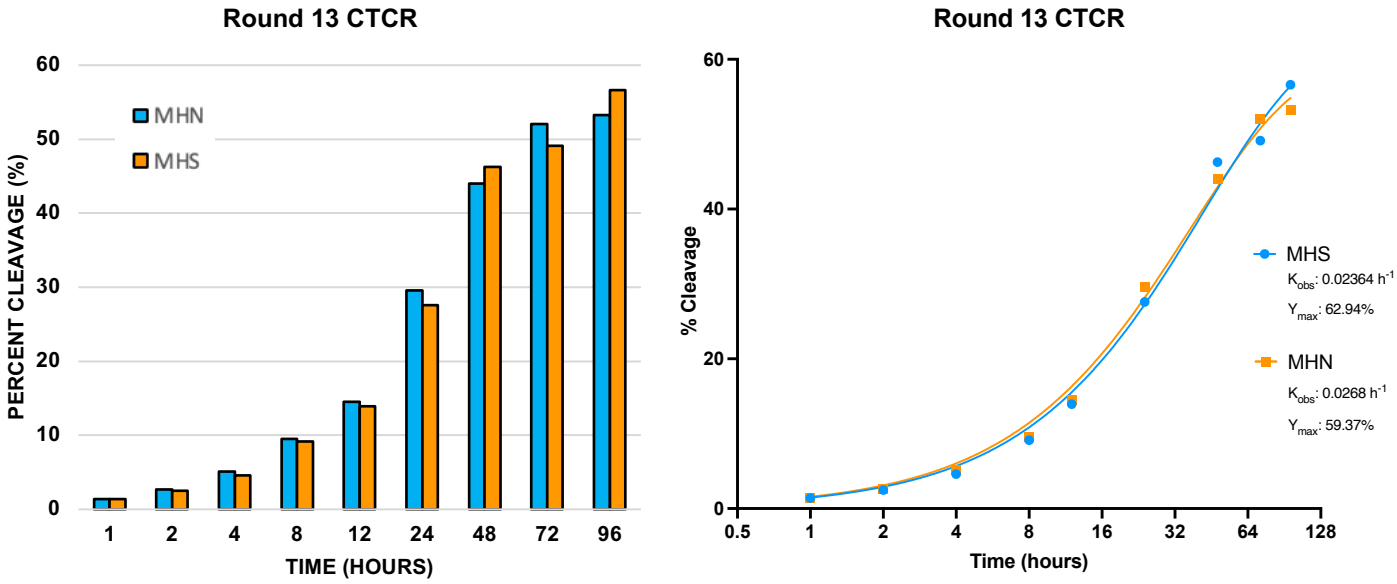


Figure 3. A. Bar graph comparison of percent cleavage value of Round 13 CTCR. **B.** Kinetic plot of Round 13 CTCR with associated rate constant and predicted plateau.

The results of both the raw percent cleavage data and the kinetic analysis show remarkably consistent activity of the DNAzyme library in MHS and MHN serum through 96 hours. In fact, a steady increase in cleavage over time is noted in both MHS and MHN serum, with comparable raw cleavage values for each time point. This observation is further validated by the kinetic data, which yielded very similar DNAzyme library rate constant values and predicted plateaus. Such consistency in activity could be attributed to performing a selection in serum from a non-acute condition, whereby a lack of active metabolic processes make it more challenging for the DNAzyme library to readily distinguish biomarker differences between the samples, and avoid

responding to common targets. This may suggest the need for additional rounds of selection. However, given the degree of percent cleavage of the DNAzyme library observed by Round 12, and the possibility of losing library diversity to non-specific nuclease degradation, we opted to stop the selection and proceed with identification of candidate DNAzymes for sensitivity and specificity analysis.

3.2.2 Identification of candidate DNAzyme sequences targeting malignant hyperthermia sensitive serum using sequencing data

Upon reviewing the results of the sequencing data, we observed a total of 213,142 classes of sequences in the Round 6 population, 307,663 in the Round 7 population, 242,271 in the Round 8 population, 217,169 in the Round 9 population, 208,756 in the Round 10 population, 134,402 in the Round 11 population, and 147,013 sequence classes in the Round 12 population. Although the total number of sequence classes appears to decrease over the course of the selection, the difference in the raw class numbers from Round 6 to Round 12 is quite low, and possibly indicative of a plateau. This would imply exhaustion of the current selection conditions, consistent with the raw cleavage data results. There does however appear to be considerable reshuffling of the ranking of top sequence classes based on their fluctuating frequency over the latter 7 rounds of selection, indicating ongoing enrichment of the library. This effect is particularly noticeable in the top four sequences at Round 12.

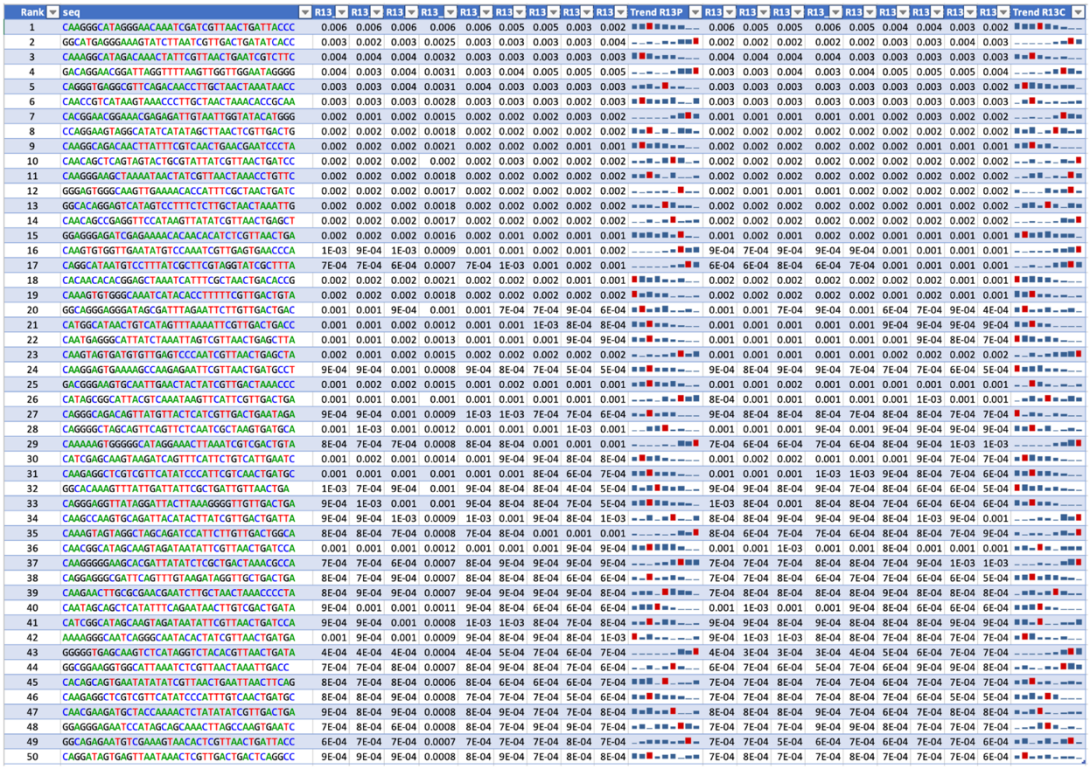
The decision to include the cleavage time course reactions in the sequencing run was driven by the need to acquire more directional input from the sequencing data (as compared to the pancreatic cancer selection). As such, the goal of sequencing individual time points was envisioned to allow

for a richer sequencing analysis beyond rudimentary/fundamental observations of library enrichment. In effect, by introducing time as a criterion of sequence sorting in addition to fraction of total library, we can identify additional specific enrichment patterns. Doing so would facilitate clustering of sequences based on their performance over time in addition to conservation of sequence alignment/identity. Furthermore, sequencing individual time points can also enable us to track sequence classes or notable motifs through time and across MHS and MHN serum, allowing us to identify an optimal time point for differentiation between serums, and highlighting sequence candidates best suited for conducting the differentiation at each time point. Figure 4 presents a summary of the top 50 sequences at Round 12, their enrichment over the prior 6 rounds of selection, and their performance in the Round 13 cleavage time course reaction.

A

Rank	seq	R6	R7	R8	R9	R10	R11	R12	Trend
1	CAAGGGCATAGGGAACAAATCGATCGTTAACTGATTACCC	0.00016617	0.000404	0.00109057	0.00187163	0.00278056	0.00407884	0.0050975	
2	GGCATGAGGGAAAGTATCTTAATCGTTGACTGATATCACC	7.93E-05	0.00012835	0.0003518	0.00060273	0.00100204	0.00173162	0.00270957	
3	CAAAGGCATAGACAAACTATTCGTTAACTGAATCGTCTTC	0.00026436	0.00031352	0.00066428	0.00120359	0.00178141	0.0023915	0.00259247	
4	GACAGGAACGGATTAGGTTTTAAGTTGGTTGGAATAGGGG		2.95E-05	6.62E-05	0.00015861	0.00050464	0.00112046	0.00249316	
5	CAGGGTGAGGCGTTCAGACAACCTTGCTAACTAAATAACC	2.27E-05	9.89E-05	0.00019659	0.00051316	0.00097312	0.00125111	0.00195807	
6	CAACCGTCATAAGTAAACCCTTGCTAACTAAACACCGCAA	0.00010197	0.00017044	0.00036835	0.000599	0.00102373	0.00143712	0.00194176	
7	CACGGAACGGAAACGAGAGATTGTAATTGGTATACATGGG	2.64E-05	9.47E-05	0.00015934	0.00032469	0.00067237	0.00102967	0.00167051	
8	CCAGGAAGTAGGCATATCATATAGCTTAACTCGTTGACTG	2.27E-05	3.37E-05	0.00012623	0.00023699	0.00043089	0.00090124	0.00165717	
9	CAAGGCAGACAACTTATTTCGTCAACTGAACGAATCCCTA	0.00016995	0.00023356	0.00048631	0.0007968	0.00109747	0.00122675	0.00154452	
10	CAACAGCTCAGTAGTACTGCGTATTATCGTTAACTGATCC	9.44E-05	0.00013467	0.00035387	0.00061206	0.00083865	0.00129983	0.00151932	
11	CAAGGGAAGCTAAAATAACTATCGTTAACTAAACCTGTTC	5.29E-05	7.57E-05	0.00016555	0.00032096	0.00051042	0.00086138	0.00130291	
12	GGGAGTGGGCAAGTTGAAAACACCATTTCGCTAACTGATC	0.00012085	0.00018096	0.00032283	0.00049636	0.00083865	0.00109611	0.00128808	
13	GGCACAGGAGTCATAGTCCTTTCTCTTGCTAACTAAATTG	3.78E-05	5.47E-05	0.00016555	0.00028737	0.00054801	0.00074181	0.0012451	
14	CAACAGCCGAGGTTCCATAAGTTATATCGTTAACTGAGCT	1.51E-05	6.31E-05	0.00015314	0.00031536	0.00059139	0.00081045	0.00115913	
15	GGAGGGAGATCGAGAAAACACAACACATCTCGTTAACTGA	6.04E-05	8.00E-05	0.00018211	0.00033029	0.0005726	0.00074845	0.0011532	
16	CAAGTGTGGTTGAATATGTCCAAATCGTTGAGTGAACCCA	0.00014729	0.00018306	0.00038698	0.00050383	0.00066947	0.00097432	0.00113393	
17	CAGGCATAATGTCCTTTATCGCTTCGTAGGTATCGCTTTA	7.55E-05	9.68E-05	0.00015107	0.00036014	0.00058416	0.00081488	0.00109391	
18	CACAACACACGGAGCTAAATCATTTCGCTAACTGACACCG	0.00011707	0.00016202	0.00032903	0.00043479	0.00057549	0.00098317	0.00107019	
19	CAAAGTGTGGGCAAATCATACACCTTTTTCGTTGACTGTA	6.42E-05	7.15E-05	0.00015727	0.00025938	0.00055958	0.00077945	0.00102276	
20	GGCAGGGAGGGATAGCGATTTAGAATTCTTGTTGACTGAC	1.51E-05	7.57E-05	0.00010761	0.00032282	0.00048584	0.00051816	0.00094865	
21	CATGGCATAACTGTCATAGTTTAAAATTCGTTGACTGACC	0.00018883	0.00015992	0.00031248	0.00046651	0.00069405	0.00083703	0.00090714	
22	CAATGAGGGCATTATCTAAATTAGTCGTTAACTGAGCTTA	1.51E-05	5.89E-05	0.00012209	0.00021833	0.00041932	0.00058459	0.00090121	
23	CAAGTAGTGATGTGTTGAGTCCCAATCGTTAACTGAGCTA	7.55E-05	8.00E-05	0.00019659	0.00032282	0.00052199	0.00078831	0.00089232	
24	CAAGGAGTGAAAAGCCAAGAGAATTCGTTAACTGATGCCT	7.18E-05	0.00011152	0.00027937	0.00037694	0.00057982	0.00079495	0.00087898	
25	GACGGGAAGTGCAATTGAACTACTATCGTTGACTAAACCC	3.40E-05	5.68E-05	0.00014693	0.00026124	0.00041065	0.00062888	0.00085378	
26	CATAGCGGCATTACGTCAAATAAGTTCATTCGTTGACTGA	2.27E-05	5.26E-05	0.00016141	0.00023139	0.00046126	0.00058238	0.00084637	
27	CAGGGCAGACAGTTATGTTACTCATCGTTGACTGAATAGA	3.78E-05	6.73E-05	0.00015314	0.00021833	0.00041065	0.00067981	0.00081524	
28	CAGGGGCTAGCAGTTCAGTTCTCAATCGCTAAGTGATGCA	1.89E-05	5.05E-05	0.00011589	0.00024258	0.00034124	0.00049823	0.00079449	
29	CAAAAAGTGGGGGCATAGGAAACTTAAATCGTCGACTGTA	3.02E-05	8.21E-05	0.0002028	0.00029857	0.00044824	0.00056687	0.00076485	
30	CATCGAGCAAGTAAGATCAGTTTCATTCTGTCATTGAATC	3.78E-05	0.00013677	0.00024833	0.00036014	0.00045403	0.0005713	0.0007604	
31	CAAGAGGCTCGTCGTTCATATCCCATTCGTCAACTGATGC	3.40E-05	8.21E-05	0.00024212	0.00032096	0.00048005	0.00070859	0.00073075	
32	GGCACAAAGTTTATTGATTATTCGCTGATTGTTAACTGA	9.82E-05	8.84E-05	0.00022763	0.00027804	0.00029208	0.00052923	0.00071593	
33	CAGGGAGGTTATAGGATTACTTAAAGGGGTTGTTGACTGA	3.02E-05	5.05E-05	0.00013037	0.00018287	0.00037161	0.00049602	0.00070556	
34	CAAGCCAAGTGCAGATTACATACTTATCGTTGACTGATTA	4.53E-05	8.21E-05	0.00018004	0.00029857	0.00037161	0.00060673	0.00069666	
35	CAAAGTAGTAGGCTAGCAGATCCATTCTTGTTGACTGGCA	1.13E-05	3.16E-05	6.83E-05	0.00015301	0.00026461	0.00043401	0.0006937	
36	CAACGGCATAGCAAGTAGATAATATTCGTTAACTGATCCA	1.89E-05	5.47E-05	0.00011796	0.00019967	0.00033257	0.00049823	0.00069073	
37	CAAGGGGGAAGCACGATTATATCTCGCTGACTAAACGCCA	6.42E-05	0.00014098	0.00022763	0.00032282	0.00043523	0.00059566	0.00066554	
38	CAGGAGGGCGATTCAGTTTGTAAGATAGGTTGCTGACTGA	3.78E-06	1.26E-05	6.83E-05	0.00013249	0.00027473	0.00043623	0.00066405	
39	CAAGAACTTGCGCGAACGAATCTTGCTAACTAAACCCCTA	1.13E-05	4.00E-05	8.28E-05	0.00016981	0.00028919	0.00044509	0.00065664	
40	CAATAGCAGCTCATATTTCAGAATAACTTGTCGACTGATA	5.66E-05	8.21E-05	0.00014693	0.00021459	0.00042366	0.00050044	0.00063293	
41	CATCGGCATAGCAAGTAGATAATATTCGTTAACTGATCCA	5.29E-05	7.15E-05	0.00012003	0.000209	0.00034703	0.00042073	0.00062107	
42	AAAAGGGCAATCAGGGCAATACACTATCGTTAACTGATGA	1.51E-05	4.63E-05	7.86E-05	0.00018474	0.00029353	0.00046723	0.00060773	
43	GGGGGTGAGCAAGTCTCATAGGTCTACACGTTAACTGATA	6.42E-05	0.00012204	0.00022763	0.00030603	0.0003933	0.00042073	0.00060773	
44	GGCGGAAGGTGGCATTAAATCTCGTTAACTAAATTGACC	1.13E-05	6.94E-05	0.00012209	0.00022952	0.00032245	0.00033658	0.00060328	
45	CACAGCAGTGAATATATATCGTTAACTGAATTAACTTCAG	3.02E-05	8.42E-05	0.00015314	0.00029297	0.00049885	0.00061116	0.00059439	
46	CAAGAGGCTCGTCGTTCATATCCCATTTGTCAACTGATGC	1.13E-05	7.36E-05	0.00011796	0.00022579	0.00038462	0.00050709	0.00057956	
47	CAACGAAGATGCTACCAAAACTCTATATATCGTTGACTGA	2.64E-05	4.42E-05	0.00010968	0.00019034	0.00031955	0.00039858	0.00057808	
48	GGAGGGAGAATCCATAGCAGCAAACTTAGCCAAGTGAATC	0.00014351	0.00018727	0.00034973	0.00046278	0.00049741	0.0005713	0.00057215	
49	GGCAGAGAATGTCGAAAGTAACACTCGTTAACTGATTACC	2.27E-05	5.05E-05	9.52E-05	0.00015861	0.00034992	0.00045616	0.00057215	
50	CAGGATAGTGAGTTAATAAACTCGTTGACTGACTCAGGCC	4.15E-05	5.89E-05	0.00012623	0.00022952	0.00034269	0.00048494	0.00057215	

B



Rank	seq	R13	R13	R13	R13	R13	R13	R13	R13	R13	R13	Trend R13P	R13	R13	R13	R13	R13	R13	R13	R13	R13	Trend R13C
1	CAAGGGCATAGGGAACAAATCGATCGTTAACTGATTACCC	0.006	0.006	0.006	0.006	0.006	0.005	0.005	0.003	0.002			0.006	0.005	0.005	0.006	0.005	0.004	0.004	0.003	0.002	
2	GGCATGAGGGAAAGTATCTTAATCGTTGACTGATATCACC	0.003	0.002	0.003	0.0025	0.003	0.003	0.003	0.003	0.004			0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.004	0.003	
3	CAAAGGCATAGACAAACTATTCGTTAACTGAATCGTCTTC	0.004	0.004	0.004	0.0032	0.003	0.003	0.003	0.003	0.003			0.004	0.004	0.004	0.004	0.003	0.003	0.003	0.003	0.003	
4	GACAGGAACGGATTAGGTTTTAAGTTGGTTGGAATAGGGG	0.004	0.003	0.004	0.0031	0.003	0.004	0.005	0.005	0.005			0.003	0.003	0.004	0.003	0.004	0.005	0.005	0.004		
5	CAGGGTGAGGCGTTCAGACAACCTTGCTAACTAAATAACC	0.003	0.003	0.004	0.0031	0.004	0.003	0.003	0.003	0.003			0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.002	
6	CAACCGTCATAAGTAAACCCTTGCTAACTAAACACCGCAA	0.003	0.003	0.003	0.0028	0.003	0.003	0.003	0.002	0.003			0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	
7	CACGGAACGGAAACGAGAGATTGTAATTGGTATACATGGG	0.002	0.001	0.002	0.0015	0.002	0.002	0.002	0.003	0.002			0.001	0.001	0.001	0.001	0.001	0.002	0.003	0.002	0.002	
8	CCAGGAAGTAGGCATATCATATAGCTTAACTCGTTGACTG	0.002	0.002	0.002	0.0018	0.002	0.002	0.002	0.002	0.002			0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	
9	CAAGGCAGACAACTTATTTCGTCAACTGAACGAATCCCTA	0.002	0.002	0.002	0.0021	0.002	0.002	0.002	0.001	0.001			0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001	
10	CAACAGCTCAGTAGTACTGCGTATTATCGTTAACTGATCC	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.002	0.002			0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	
11	CAAGGGAAGCTAAAATAACTATCGTTAACTAAACCTGTTC	0.002	0.002	0.002	0.0018	0.002	0.002	0.002	0.002	0.002			0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	
12	GGGAGTGGGCAAGTTGAAAACACCATTTCGCTAACTGATC	0.002	0.002	0.002	0.0017	0.002	0.002	0.002	0.002	0.002			0.002	0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.002	
13	GGCACAGGAGTCATAGTCCTTTCTCTTGCTAACTAAATTG	0.002	0.002	0.002	0.0018	0.002	0.002	0.002	0.002	0.002			0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.002	0.002	
14	CAACAGCCGAGGTTCCATAAGTTATATCGTTAACTGAGCT	0.002	0.002	0.002	0.0017	0.002	0.002	0.002	0.002	0.002			0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	
15	GGAGGGAGATCGAGAAAACACAACACATCTCGTTAACTGA	0.002	0.002	0.002	0.0016	0.002	0.001	0.002	0.001	0.001			0.001	0.002	0.001	0.002	0.002	0.001	0.001	0.001	0.001	
16	CAAGTGTGGTTGAATATGTCCAAATCGTTGAGTGAACCCA	1E-03	9E-04	1E-03	0.0009	0.001	0.001	0.002	0.001	0.002			9E-04	7E-04	9E-04	9E-04	9E-04	0.001	0.001	0.001	0.001	
17	CAGGCATAATGTCCTTTATCGCTTCGTAGGTATCGCTTTA	7E-04	7E-04	6E-04	0.0007	7E-04	1E-03	0.001	0.002	0.001			6E-04	6E-04	8E-04	6E-04	7E-04	0.001	0.001	0.001	0.001	
18	CACAACACACGGAGCTAAATCATTTCGCTAACTGACACCG	0.002	0.002	0.002	0.0021	0.002	0.002	0.002	0.001	0.001			0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001	
19	CAAAGTGTGGGCAAATCATACACCTTTTTCGTTGACTGTA	0.002	0.002	0.002	0.0018	0.002	0.002	0.002	0.002	0.002			0.002	0.002	0.002	0.002	0.002	0.001	0.002	0.001	0.001	
20	GGCAGGGAGGGATAGCGATTTAGAATTCTTGTTGACTGAC	0.001	0.001	9E-04	0.001	0.001	7E-04	7E-04	9E-04	6E-04			0.001	0.001	7E-04	9E-04	0.001	6E-04	7E-04	9E-04	4E-04	
21	CATGGCATAACTGTCATAGTTTAAAATTCGTTGACTGACC	0.001	0.001	0.002	0.0012	0.001	0.001	1E-03	8E-04	8E-04			0.001	0.001	0.001	0.001	0.001	9E-04	9E-04	9E-04	9E-04	
22	CAATGAGGGCATTATCTAAATTAGTCGTTAACTGAGCTTA	0.001	0.001	0.002	0.0013	0.001	0.001	0.001	9E-04	9E-04			0.001	0.001	0.001	0.001	0.001	0.001	9E-04	8E-04	7E-04	
23	CAAGTAGTGATGTGTTGAGTCCCAATCGTTAACTGAGCTA	0.002	0.001	0.002	0.0015	0.002	0.002	0.002	0.002	0.002			0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002	
24	CAAGGAGTGAAAAGCCAAGAGAATTCGTTAACTGATGCCT	9E-04	9E-04	0.001	0.0008	9E-04	8E-04	7E-04	5E-04	5E-04			9E-04	8E-04	9E-04	9E-04	7E-04	7E-04	6E-04	5E-04		
25	GACGGGAAGTGCAATTGAACTACTATCGTTGACTAAACCC	0.001	0.002	0.002	0.0015	0.001	0.002	0.001	0.001	0.001			0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	
26	CATAGCGGCATTACGTCAAATAAGTTCATTCGTTGACTGA	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001			8E-04	0.001	0.001	0.001	0.001	0.001	1E-03	0.001	0.001	
27	CAGGGCAGACAGTTATGTTACTCATCGTTGACTGAATAGA	9E-04	9E-04	0.001	0.0009	1E-03	1E-03	7E-04	7E-04	6E-04			9E-04	8E-04	8E-04	8E-04	7E-04	8E-04	8E-04	7E-04	7E-04	
28	CAGGGGCTAGCAGTTCAGTTCTCAATCGCTAAGTGATGCA	0.001	1E-03	0.001	0.0012	0.001	0.001	0.001	1E-03	0.001			0.001	0.001	0.001	9E-04	0.001	9E-04	9E-04	9E-04	9E-04	
29	CAAAAAGTGGGGGCATAGGAAACTTAAATCGTCGACTGTA	8E-04	7E-04	7E-04	0.0008	8E-04	8E-04	0.001	0.001	0.001			7E-04	6E-04	6E-04	7E-04	6E-04	8E-04	9E-04	1E-03	1E-03	
30	CATCGAGCAAGTAAGATCAGTTTCATTCTGTCATTGAATC	0.001	0.002	0.001	0.0014	0.001	9E-04	9E-04	8E-04	8E-04			0.001	0.002	0.002	0.001	0.001	0.001	9E-04	7E-04	7E-04	
31	CAAGAGGCTCGTCGTTCATATCCCATTCGTCAACTGATGC	0.001	0.001	0.001	0.001	0.001	0.001	8E-04	6E-04	7E-04			0.001	0.001	0.001	1E-03	1E-03	9E-04	8E-04	7E-04	6E-04	
32	GGCACAAAGTTTATTGATTATTCGCTGATTGTTAACTGA	1E-03	7E-04	9E-04	0.001	9E-04	8E-04	8E-04	4E-04	5E-04			9E-04	9E-04	8E-04	9E-04	7E-04	8E-04	6E-04	6E-04	5E-04	
33	CAGGGAGGTTATAGGATTACTTAAAGGGGTTGTTGACTGA	9E-04	1E-03	0.001	0.001	9E-04	8E-04	8E-04	7E-04	5E-04			1E-03	8E-04	0.001	8E-04	8E-04	7E-04	6E-04	6E-04	6E-04	
34	CAAGCCAAGTGCAGATTACATACTTATCGTTGACTGATTA	9E-04	9E-04	1E-03	0.0009	1E-03	0.001	9E-04	8E-04	1E-03			8E-04	8E-04	9E-04	9E-04	9E-04	8E-04	1E-03	9E-04	0.001	
35	CAAAGTAGTAGGCTAGCAGATCCATTCTTGTTGACTGGCA	8E-04	8E-04	7E-04	0.0008	7E-04	8E-04	0.001	0.001	0.001			8E-04	6E-04	7E-04	7E-04	6E-04	8E-04	8E-04	7E-04	9E-04	
36	CAACGGCATAGCAAGTAGATAATATTCGTTAACTGATCCA	0.001	0.001	0.001	0.0012	0.001	0.001	0.001	9E-04	9E-04			0.001	0.001	1E-03	0.001	0.001	8E-04	0.001	0.001	0.001	
37	CAAGGGGGAAGCACGATTATATCTCGCTGACTAAACGCCA	7E-04	7E-04	6E-04	0.0007	8E-04	9E-04	9E-04	9E-04	9E-04			7E-04	7E-04	7E-04	8E-04	8E-04	7E-04	9E-04	1E-03	1E-03	
38	CAGGAGGGCGATTCAGTTTGTAAGATAGGTTGCTGACTGA	8E-04	7E-04	9E-04	0.0007	8E-04	8E-04	8E-04	6E-04	6E-04			7E-04	7E-04	8E-04	6E-04	7E-04	7E-04	6E-04	6E-04	5E-04	
39	CAAGAACTTGCGCGAACGAATCTTGCTAACTAAACCCCTA	8E-04	9E-04	9E-04	0.0007	9E-04	9E-04	9E-04	9E-04	8E-04			7E-04	8E-04	8E-04	8E-04	9E-04	8E-04	8E-04	8E-04	8E-04	
40	CAATAGCAGCTCATATTTCAGAATAACTTGTCGACTGATA	9E-04	0.001	0.001	0.0011	9E-04	8E-04	6E-04	6E-04	6E-04			0.001	1E-03	0.001	0.001	9E-04	8E-04	6E-04	6E-04	6E-04	
41	CATCGGCATAGCAAGTAGATAATATTCGTTAACTGATCCA	9E-04	9E-04	0.001	0.0008	1E-03	1E-03	8E-04	7E-04	8E-04			9E-04	9E-04	8E-04	9E-04	8E-04	8E-04	8E-04	9E-04	9E-04	
42	AAAAGGGCAATCAGGGCAATACACTATCGTTAACTGATGA	0.001	9E-04	0.001	0.0009	9E-04	8E-04	9E-04	8E-04	1E-03			9E-04	1E-03	1E-03	8E-04	8E-04	7E-04	8E-04	7E-04	7E-04	
43	GGGGGTGAGCAAGTCTCATAGGTCTACACGTTAACTGATA	4E-04	4E-04	4E-04	0.0004	4E-04	5E-04	7E-04	6E-04	7E-04			4E-04	3E-04	3E-04	3E-04	4E-04	5E-04	6E-04	7E-04	7E-04	
44	GGCGGAAGGTGGCATTAAATCTCGTTAACTAAATTGACC	7E-04	7E-04	8E-04	0.0007	8E-04	9E-04	8E-04	7E-04	6E-04			6E-04	7E-04	6E-04	5E-04	7E-04	6E-04	9E-04	9E-04	8E-04	
45	CACAGCAGTGAATATATATCGTTAACTGAATTAACTTCAG	8E-04	7E-04	8E-04	0.0006	8E-04	6E-04	6E-04	6E-04	7E-04			7E-04	7E-04	8E-04	7E-04	7E-04	6E-04	7E-04	7E-04	7E-04	
46	CAAGAGGCTCGTCGTTCATATCCCATTTGTCAACTGATGC	8E-04	8E-04	9E-04	0.0008	7E-04	7E-04	7E-04	5E-04	6E-04			7E-04	7E-04	8E-04	7E-04	8E-04	7E-04	6E-04	5E-04	5E-04	
47	CAACGAAGATGCTACCAAAACTCTATATATCGTTGACTGA	9E-04	8E-04	9E-04	0.0008	8E-04	9E-04	7E-04	7E-04	8E-04			8E-04	8E-04	8E-04	9E-04	7E-04	6E-04	7E-04	8E-04	8E-04	
48	GGAGGGAGAATCCATAGCAGCAAACTTAGCCAAGTGAATC	7E-04	8E-04	6E-04	0.0008	8E-04	7E-04	9E-04	9E-04	7E-04			7E-04	7E-04	7E-04	9E-04	9E-04	8E-04	7E-04	7E-04	7E-04	
49	GGCAGAGAATGTCGAAAGTAACACTCGTTAACTGATTACC	6E-04	7E-04	7E-04	0.0007	7E-04	7E-04	8E-04	7E-04				7E-04	7E-04	5E-04	6E-04	7E-04	6E-04	6E-04	7E-04	6E-04	
50	CAGGATAGTGAGTTAATAAACTCGTTGACTGACTCAGGCC	9E-04	9E-04	9E-04	0.0008	8E-04	9E-04	7E-04	8E-04	8E-04			7E-04	8E-04	7E-04	7E-04	8E-04	7E-04	7E-04	7E-04	6E-04	

Figure 4. A. Top 50 sequences through Rounds 6-12, in order of frequency at Round 12. **B.** Performance of top 50 R12 sequences in the CTCR (left = positive selection serum, right = counter selection serum).

Interestingly, Figure 4A shows all 50 top sequences at Round 12 are gradually enriching over the course of the selection, with the exception of rank 45, which appears to peak in Round 11. In the absence of any clearly distinct trends in enrichment that would garner curiosity, it appears the conventional approach of testing the top four ranked sequences in the last round of selection is a reasonable approach in our study as well. Remarkably, when comparing their performance in the Round 13 cleavage time course reaction, the top 4 sequences behave quite differently in both MHS and MHN serum. Rank 1 and 3 show peak cleavage activity earlier in the time courses, and faster in MHS than MHN serum. Conversely, Rank 2 and 4 reach peak cleavage at the end of the 96 hour time course, and later in MHS serum than MHN. Given that there do not appear to be any time point-peak cleavage overlap in the two types of serum for all top 4 sequences, we can infer they

hold positive/counter selection serum differentiating potential, and are reassured to proceed with their testing. In addition to ascertaining whether any of these candidates can proceed into diagnostic assay construction, testing these DNAzyme candidates will also allow us to determine whether the time point-associated cleavage patterns emerging from sequencing the cleavage time course reactions correlate with the actual performances of the candidates. If so, we expect the sensitivity and specificity analysis to yield up to two early-resulting and 2 late-resulting DNAzyme probes for MHS detection.

Lastly, sequencing analysis shows that among the top 50 sequences, the 4 hour time point often demonstrated peak DNAzyme activity in MHS serum, correlating with the positive selection incubation time. This observation reflects the concept of SELEX and *in vitro* evolution of the DNAzyme library through implementation of selection pressures, in this case teaching the library to bind to a target of interest within the allocated 4 hour incubation time. Consequently, careful consideration will be given to the cleavage percentage values at the 4 hour time point for each of the four candidate sequence tested, and compared across MHS and MHN serum. The DNAzyme candidates selected for sensitivity and specificity analysis, their predicted folding structure, and their associated sequencing summary are denoted in Figure 5.

A

	Round 6		Round 12	
Total Classes	213,142		147,013	
Sequences	**Rank**	**Frequency (%)**	**Rank**	**Frequency (%)**
12-1	5	0.0166	1	0.5097
12-2	39	0.0079	2	0.2710
12-3	1	0.0264	3	0.2592
12-4	25,325	Undetectable	4	0.2493

B

12-1 5'-CTATGAACTGAC-Q-R-F-GACCTCACTACCAAGCGCACCGTAGCAGATGACCAAGGGCATAGGGAACAAATCGATCGTTAACTGATTACCCTCATTGTTTCAGTGGCGAGCA

12-2 5'-CTATGAACTGAC-Q-R-F-GACCTCACTACCAAGCGCACCGTAGCAGATGACGGCATGAGGGAAAGTATCTTAATCGTTGACTGATATCACCTCATTGTTTCAGTGGCGAGCA

12-3 5'-CTATGAACTGAC-Q-R-F-GACCTCACTACCAAGCGCACCGTAGCAGATGACCAAAGGCATAGACAAACTATTCGTTAACTGAATCGTCTTCTCATTGTTTCAGTGGCGAGCA

12-4 5'-CTATGAACTGAC-Q-R-F-GACCTCACTACCAAGCGCACCGTAGCAGATGACGACAGGAACGGATTAGGTTTTAAGTTGGTTGGAATAGGGGTCATTGTTTCAGTGGCGAGCA

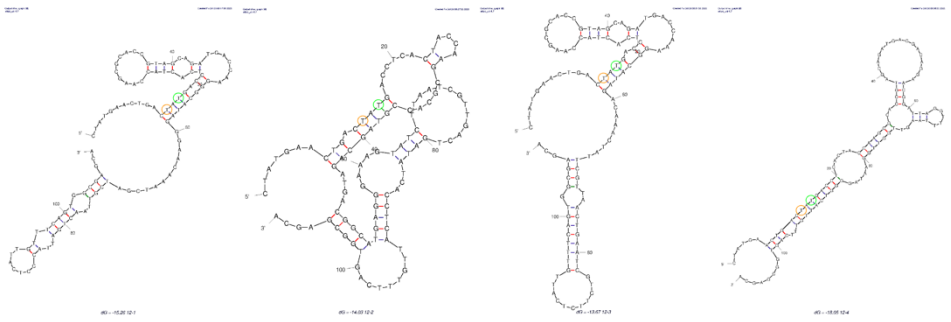


Figure 5. A Summary of sequencing results. Undetectable refers to a value < 3.78E-06. **B.** Sequence identities of candidate DNAzymes and their predicted folding structures. The fluorophore (green) and quencher (orange) locations along the sequences are encircled.

With the exception of Rank 1 and 3, the predicted folding structures of the DNAzyme candidates show remarkable differences, which could be indicative of different target interactions. Another notable observation is the folding structure around the fluorophore and quencher moieties, particularly when compared to the pancreatic cancer selection DNAzymes' predicted folding structures. The MH selection DNAzyme candidates' fluorophore and quencher are positioned along a linear structural motif compared to a loop motif in the pancreatic cancer selection

DNAzymes. This observation could see the MH selection DNAzyme candidates being opened up to new target interactions, avoiding the likelihood of both selections yielding DNAzymes that recognize a common non-specific serum biomarker.

3.2.3 Sensitivity and Specificity Analysis

Once again, the four DNAzymes candidates were first screened for cross-reactivity to selection buffer/self-cleavage before proceeding with sensitivity and specificity validation. The results of this analysis showed sub-1% cleavage for all four sequences through 24 hours, indicating they are optimized to proceed to sensitivity and specificity analysis with individual patient serum samples.

Sensitivity

The sensitivity analysis was designed to gauge the ability of each of the top 4 DNAzyme sequences to converge on the positive CHCT diagnosis of MH susceptibility. Consequently, each of the four DNAzyme candidates was incubated with individual patient serum from 45 MHS patients, and underwent cleavage time course reactions over 24 hours. The results of the sensitivity analysis are outlined in Figure 6. The 10 samples used to create the positive selection serum pool for the *in vitro* selection protocol are ordered at the top of each sensitivity results table. Given that background cleavage of DNAzymes in buffer-only reactions peaked at 0.8%, we can infer that any cleavage over 1% is attributable to serum components. Therefore, all reactions with peak cleavage >1% are considered positive results.

12-1

Sample ID	TIME (hours) 1.5	3	4.5	8	12	24
1504	0.83	1.47	3.41	8.80	10.75	25.65
1511	1.90	2.24	3.40	8.84	12.30	26.57
1518	0.73	1.29	3.14	7.86	11.36	23.59
1543	0.47	1.19	3.31	7.18	10.58	22.92
1553	1.68	2.26	4.08	9.86	14.16	31.24
1562	1.90	1.99	3.05	5.90	6.52	16.65
1570	1.04	1.69	3.40	7.67	10.25	26.28
1571	3.15	4.89	9.36	21.46	27.30	43.81
1578	2.12	3.83	6.84	14.94	20.62	35.10
1579	1.47	2.02	3.14	8.55	10.57	20.37
1509	2.05	2.41	3.59	5.47	6.44	12.30
1523	12.58	15.26	15.75	26.45	35.22	51.32
1529	0.93	1.60	1.74	1.80	1.89	2.59
1532	1.01	1.52	2.01	3.43	5.13	10.70
1533	0.56	0.39	0.90	0.63	1.05	0.88
1547	1.26	1.31	1.48	1.55	1.74	2.13
1573	1.68	2.08	1.45	1.53	1.63	1.73
1575	1.44	1.37	1.56	1.57	1.94	2.44
1577	1.38	1.78	1.43	2.20	2.42	2.59
1580	0.65	1.05	2.09	1.97	2.02	2.28
1581	0.53	0.60	1.07	1.47	1.66	1.90
1583	1.23	1.57	1.98	2.21	2.46	3.14
1585	0.46	0.90	1.37	1.47	1.41	2.01
1594	-0.02	0.29	0.22	0.41	0.42	0.41
1595	1.15	1.34	1.60	1.71	2.20	1.89
1597	4.60	5.43	5.34	6.03	5.98	6.50
1601	0.31	0.38	0.77	1.26	1.24	1.57
1607	1.10	1.05	1.47	1.33	1.52	1.69
1612	0.76	1.74	1.99	2.89	3.36	4.86
1613	0.94	1.13	1.46	1.72	1.80	1.99
1614	3.35	3.45	3.11	2.83	2.93	2.70
1619	0.21	0.55	0.00	0.73	0.75	0.70
1622	0.54	1.42	1.39	1.09	1.32	0.27
1623	2.03	1.75	2.15	2.02	1.82	1.99
1637	2.76	2.80	3.61	3.24	3.34	3.32
1642	0.24	0.58	0.85	0.61	0.75	0.97
1649	1.61	1.87	2.03	2.29	2.34	2.47
1650	1.68	1.45	1.45	1.86	1.94	2.33
1653	1.92	1.83	2.44	2.70	3.19	3.33
1654	2.97	3.67	2.86	3.44	3.19	2.94
1655	2.78	2.24	2.15	2.28	2.79	3.32
1664	0.85	1.38	1.40	1.91	2.55	2.61
1671	1.34	1.06	1.56	1.58	1.85	1.84
1676	2.32	2.39	2.28	3.11	3.14	3.39

12-2

Sample ID	TIME (hours) 1.5	3	4.5	8	12	24
1504	1.04	1.47	2.37	5.04	12.16	18.49
1511	1.46	2.30	3.42	6.26	14.30	20.75
1518	0.91	1.20	1.75	3.50	8.88	15.63
1543	1.51	2.33	3.31	5.49	13.22	19.45
1553	1.93	3.19	5.23	12.09	28.78	39.61
1562	1.23	1.81	2.46	3.98	8.48	12.81
1570	1.47	2.29	3.81	5.98	14.19	21.73
1571	2.35	4.73	7.19	13.47	30.06	39.53
1578	2.04	3.87	7.03	11.43	22.73	30.65
1579	1.81	2.57	3.83	7.00	12.16	16.36
1509	0.89	1.78	2.28	2.72	4.77	7.08
1523	0.88	1.29	2.17	4.53	6.19	7.07
1529	0.41	0.52	0.53	0.62	0.64	0.97
1532	0.42	0.98	1.44	2.52	3.45	4.49
1533	0.93	0.91	1.06	1.06	1.36	1.49
1547	0.54	0.82	0.74	0.96	1.25	1.58
1573	1.09	1.32	1.55	1.48	1.74	1.96
1575	1.62	1.48	1.83	1.56	2.08	2.06
1577	0.79	0.97	1.18	1.18	1.64	1.95
1580	0.91	1.09	1.31	1.31	1.77	2.11
1581	1.12	1.07	1.27	1.55	1.97	2.21
1583	1.12	1.25	1.25	1.28	1.37	1.43
1585	0.45	0.58	0.60	0.69	0.75	1.00
1594	0.30	0.44	0.54	0.34	0.56	0.79
1595	0.77	1.12	0.94	1.11	1.40	1.55
1597	1.38	1.26	1.39	1.47	1.68	1.91
1601	0.55	0.57	0.97	1.01	1.44	1.67
1607	1.12	1.10	1.28	1.57	1.76	1.78
1612	1.07	1.08	1.24	1.27	1.54	1.63
1613	0.66	0.88	0.93	1.25	1.53	1.61
1614	1.11	1.08	1.37	1.24	1.36	1.46
1619	1.71	1.12	1.21	1.08	1.38	1.17
1622	0.61	0.76	1.02	1.12	1.48	1.52
1623	0.25	0.34	0.39	0.40	0.67	0.67
1637	0.81	0.90	1.11	1.28	1.57	1.61
1642	0.86	1.11	1.25	1.39	1.85	2.11
1649	0.95	0.99	1.01	1.01	1.43	1.72
1650	0.75	0.71	0.89	1.01	1.48	1.78
1653	0.85	0.85	0.73	1.08	1.62	1.69
1654	1.55	1.73	1.61	1.64	2.01	1.86
1655	0.10	0.39	0.53	1.17	1.27	1.67
1664	1.35	1.60	1.65	1.61	2.06	2.38
1671	0.67	0.75	0.87	1.09	1.60	1.62
1676	1.11	0.99	1.27	1.44	1.78	2.07

12-3

Sample ID	TIME (hours) 1.5	3	4.5	8	12	24
1504	1.88	2.21	2.52	4.53	6.58	17.51
1511	0.87	1.56	2.81	4.93	8.48	20.19
1518	1.74	2.32	3.03	5.28	8.08	20.18
1543	1.53	1.97	2.65	3.98	5.94	15.13
1553	1.83	3.14	5.00	10.20	16.85	37.98
1562	1.48	2.02	2.96	4.88	6.81	15.38
1570	1.51	2.19	3.16	5.90	9.45	22.42
1571	3.11	4.89	7.67	14.10	22.45	41.51
1578	2.20	3.79	5.71	10.30	18.22	31.28
1579	1.30	2.23	2.83	5.19	7.82	16.39
1509	0.86	1.51	1.94	2.89	3.60	7.02
1523	0.65	1.29	1.90	3.03	3.85	6.97
1529	1.06	1.08	1.11	1.32	1.44	1.59
1532	0.97	1.09	1.62	2.17	3.04	5.06
1533	0.48	0.55	0.68	0.74	1.04	1.46
1547	0.43	0.41	0.27	0.78	0.98	1.31
1573	0.93	0.82	0.71	1.00	1.08	1.25
1575	0.42	0.45	0.68	0.78	0.89	1.42
1577	0.76	0.92	1.14	1.21	1.28	2.08
1580	1.27	1.17	1.25	1.53	1.41	1.86
1581	0.40	0.57	0.67	0.70	0.90	1.24
1583	0.68	1.12	1.35	1.39	1.60	1.91
1585	0.45	0.59	0.61	0.78	1.14	1.58
1594	0.60	0.71	0.84	1.21	1.21	2.15
1595	0.87	0.82	1.07	1.05	1.10	1.37
1597	0.91	0.97	1.01	1.17	1.29	1.50
1601	1.11	1.21	1.24	1.63	1.75	2.17
1607	0.91	0.89	0.93	1.13	1.19	1.53
1612	-0.66	-0.22	-0.11	0.09	0.20	0.23
1613	1.23	1.62	1.53	1.68	1.77	2.37
1614	1.41	1.30	1.42	1.59	1.63	2.12
1619	0.79	0.74	0.80	1.22	1.26	1.33
1622	0.60	0.63	0.52	0.79	1.20	1.47
1623	0.53	0.84	0.96	1.06	1.13	1.11
1637	-0.33	-0.43	-0.31	0.01	0.33	0.57
1642	0.73	0.92	1.22	1.52	1.58	2.22
1649	0.92	1.07	1.17	1.26	1.31	1.80
1650	0.61	0.69	0.49	0.95	1.36	1.05
1653	0.98	1.13	1.43	1.35	1.53	2.08
1654	1.10	1.14	1.42	1.60	1.65	1.99
1655	1.57	1.86	1.86	2.07	1.87	2.66
1664	1.50	1.89	2.12	2.29	2.49	3.84
1671	0.92	1.28	1.19	1.35	1.42	1.75
1676	1.81	2.19	1.82	2.49	2.41	3.02

12-4

Sample ID	TIME (hours) 1.5	3	4.5	8	12	24
1504	0.39	0.51	0.54	0.70	0.93	1.45
1511	0.38	0.46	0.66	0.79	0.99	1.68
1518	0.57	0.76	0.95	1.01	1.41	2.09
1543	1.06	1.44	1.37	1.80	2.69	3.20
1553	1.01	1.68	1.88	2.61	3.57	5.36
1562	1.12	1.29	1.48	1.45	1.95	3.05
1570	0.92	1.77	1.06	2.06	1.87	2.83
1571	0.92	1.37	2.10	2.04	3.44	6.65
1578	0.72	1.12	1.61	2.17	2.65	4.60
1579	0.75	0.77	0.84	1.34	1.58	2.70
1509	0.20	0.28	0.33	0.56	0.63	1.06
1523	0.41	0.31	0.35	0.18	0.35	0.55
1529	0.99	1.16	1.11	1.26	1.96	1.20
1532	1.21	1.05	1.32	1.64	1.65	1.38
1533	0.41	0.55	0.39	0.63	0.56	0.60
1547	0.41	0.48	0.56	0.53	0.62	0.63
1573	0.95	1.04	1.08	1.11	1.09	1.12
1575	1.08	1.36	1.38	1.34	1.22	1.32
1577	0.57	0.63	0.72	0.73	0.89	0.85
1580	0.56	1.33	0.66	1.20	0.86	0.85
1581	0.47	0.59	0.64	0.61	0.62	0.75
1583	0.70	0.79	0.67	0.83	0.93	0.80
1585	1.00	0.93	0.67	0.74	0.76	0.76
1594	0.47	0.73	0.72	0.81	0.76	0.80
1595	0.63	0.72	0.85	1.05	1.07	1.15
1597	0.34	0.33	0.51	0.44	0.52	0.48
1601	0.48	0.45	0.52	0.65	0.75	0.63
1607	1.27	1.23	1.19	1.15	1.01	1.20
1612	0.54	0.58	1.94	0.62	0.73	0.64
1613	0.28	0.27	0.14	0.36	0.42	0.40
1614	0.77	0.66	0.86	1.17	1.29	1.26
1619	0.46	0.36	0.46	0.18	0.23	0.54
1622	0.39	0.40	0.36	0.48	0.57	0.61
1623	0.75	0.83	0.95	1.07	0.87	1.16
1637	0.37	0.48	0.31	0.39	0.57	0.55
1642	0.63	0.73	0.86	0.85	0.68	1.07
1649	0.93	1.06	0.97	1.11	1.40	1.17
1650	0.73	0.91	1.11	1.02	1.19	1.57
1653	0.95	1.07	1.25	1.19	1.24	1.66
1654	0.75	0.90	0.78	0.95	1.04	1.01
1655	0.49	0.69	0.67	0.80	0.99	1.00
1664	0.62	0.68	0.75	0.76	0.81	0.94
1671	0.93	1.08	1.25	1.41	1.50	1.75
1676	0.95	0.82	0.84	0.86	0.91	1.06

Figure 6. Sensitivity results of top 4 DNAzyme candidates expressed as percent cleavage values.

Figure 6 shows that all DNAzyme candidates' peak cleavage occurs at 24 hours in virtually all individual patient samples, with a few exceptions largely attributed to background noise. Furthermore, according raw percent cleavage values, 12-1 and 12-3 appear to be most responsive to MHS serum at 24 hours, followed by 12-2 and 12-4. One distinctive result is seen in Patient sample 1523, whereby 12-1 in particular appears to cleave in especially high amounts and have the strongest effect. However upon reviewing the gel image, the percent value is falsely elevated secondary to large non-specific degradation of the uncleaved probe (at seemingly higher rate in), heavily skewing the percent cleavage calculation. When contextualized with the other new

samples, the degree of cleavage appears to be on par with the other new patient samples tested. Figure 6 furthermore demonstrates consistently higher/preferential DNAzyme cleavage of all 4 candidates in the patient samples used for performing the *in vitro* selection. This may suggest overfitting of the library to the initial 10 samples or could signify a true difference in the serum components from the initial samples to the new ones received for the validation study. However, clinical correlation of these samples with their specific results in the CHCT did not account for this discrepancy.

The sensitivity values of each DNAzyme candidate were compiled based on their result convergence with the CHCT. The equation employed to calculate the sensitivity[8] of each DNAzyme candidate, and the sensitivity results at 4 and 24 hours are compiled in Figure 7.

A

$$Sensitivity = \frac{True\ Positive}{(True\ Positive + False\ Negative)} \times 100$$

B

Sequence	24-hour sensitivity	4-hour sensitivity
12-1	89%	89%
12-2	89%	73%
12-3	95%	68%
12-4	59%	34%

Figure 7. A. Sensitivity calculation formula. **B.** Sensitivity summary of top 4 DNAzyme candidates at 24 hour and 4 hour detection time limit.

Impressively, 12-1 shows no difference in sensitivity when shrinking the detection time from 24 hours to 4 hours, unlike the remaining sequences which all show a sizeable decrease in sensitivity. Having a shorter detection time is undoubtedly preferential, not only for expeditious resulting of a potential future diagnostic assay, but even more importantly for limiting the degree of non-specific interactions occurring between the DNAzyme probe and serum components, and ultimately decreasing the chance of false positive results. In short, the results of the sensitivity analysis point to 12-1 being the front-runner candidate for MHS detection, given its comparatively high raw cleavage percentages, and high/unwavering sensitivity values.

Specificity

Similar to the design of the sensitivity analysis, the specificity analysis was designed to gauge the ability of each of the top 4 DNAzyme sequences to converge on the negative CHCT diagnosis of MH susceptibility. Once again, each of the four DNAzyme candidates was incubated with individual patient serum from 45 MHN patients, and underwent cleavage time course reactions over 24 hours. The results of the sensitivity analysis are outlined in Figure 8. The 10 samples used to create the counter selection serum pool for the *in vitro* selection protocol are ordered at the top of each specificity results table.

12-1

Sample ID	TIME (hours) 1.5	3	4.5	8	12	24
1535	2.02	2.91	4.66	8.15	12.38	20.41
1536	1.45	2.13	2.93	6.19	9.57	17.48
1538	1.11	1.62	2.46	4.56	7.26	18.98
1539	3.65	7.16	12.99	24.16	34.37	47.14
1545	1.92	3.48	6.25	14.21	24.03	43.37
1564	1.25	2.09	4.49	9.45	15.43	27.26
1572	2.59	3.33	5.30	8.87	14.32	24.37
1587	2.46	3.33	5.55	9.55	13.64	23.94
1590	2.13	3.97	7.03	12.91	21.90	31.91
1592	3.88	6.33	10.54	17.22	23.73	38.77
1512	0.89	1.22	1.94	2.34	3.52	6.19
1519	2.70	3.00	3.33	4.64	5.27	8.12
1521	1.72	2.08	2.47	3.33	4.08	6.78
1524	0.58	0.78	0.86	1.19	1.90	2.92
1527	2.48	1.54	2.27	2.67	3.88	6.74
1530	0.49	0.42	0.41	0.45	0.44	0.66
1548	0.29	0.38	0.28	0.46	0.96	0.82
1550	0.13	0.14	0.03	0.13	0.28	0.43
1589	0.28	0.38	0.59	0.56	0.50	0.40
1606	0.48	0.51	0.69	0.71	0.83	1.03
1610	0.63	0.71	0.97	1.32	0.96	1.64
1615	0.67	1.77	0.85	1.29	1.63	2.07
1624	0.31	0.25	0.40	0.41	0.34	0.89
1626	2.70	1.14	1.35	2.08	1.93	3.10
1629	0.77	0.93	0.99	0.78	0.96	1.10
1632	1.15	1.16	1.44	1.17	1.42	2.01
1638	0.82	0.89	0.88	1.01	1.00	1.82
1641	0.30	0.35	0.49	0.59	0.82	1.10
1643	1.59	1.63	1.83	1.98	2.04	2.50
1645	-0.43	-0.05	0.28	0.41	0.64	0.80
1646	0.28	0.19	0.26	0.15	0.32	0.44
1652	0.94	1.01	0.71	1.37	1.18	1.07
1656	1.05	1.20	1.40	1.24	1.67	1.91
1657	0.80	1.02	1.17	1.26	1.27	1.63
1658	1.11	1.19	1.33	1.36	1.62	1.62
1659	0.80	0.71	0.72	0.88	0.84	1.06
1661	1.23	1.43	1.44	1.77	1.75	1.94
1662	1.06	0.98	1.05	1.17	1.26	1.64
1665	0.95	0.98	1.01	1.06	1.01	1.49
1666	0.57	0.58	0.88	0.90	0.97	1.39
1667	1.49	1.47	1.45	1.62	1.73	2.18
1670	1.29	1.22	1.36	1.64	1.63	1.93
1673	0.81	0.70	1.01	1.29	1.44	1.66
1674	0.50	0.54	0.56	0.62	0.68	0.78
1675	0.50	0.53	0.71	0.89	1.11	1.56

12-2

Sample ID	TIME (hours) 1.5	3	4.5	8	12	24
1535	0.86	1.33	2.47	4.93	11.02	15.33
1536	0.97	1.39	1.89	3.35	8.69	11.89
1538	0.57	0.91	1.63	2.55	6.22	8.81
1539	4.48	8.66	13.82	23.79	44.86	50.21
1545	2.07	3.45	5.66	11.72	24.80	31.84
1564	2.45	3.68	5.15	7.70	12.47	14.26
1572	1.95	2.44	2.97	5.23	11.47	15.23
1587	1.87	2.36	3.39	6.55	11.54	14.01
1590	1.53	2.82	4.78	9.72	22.73	30.78
1592	2.10	3.95	6.46	11.14	26.61	36.19
1512	0.73	1.11	1.42	2.38	3.72	5.90
1519	0.97	1.14	1.51	2.31	3.56	4.59
1521	0.67	1.04	1.41	1.75	3.55	4.71
1524	1.16	1.08	1.53	1.66	2.16	2.52
1527	1.09	1.37	1.57	2.00	3.75	5.48
1530	0.35	0.50	0.58	0.69	0.96	0.92
1548	0.56	0.80	0.81	0.82	0.91	1.16
1550	0.51	0.58	0.69	0.81	0.99	1.32
1589	0.45	0.52	0.41	0.52	0.53	0.61
1606	0.39	0.39	0.39	0.39	0.41	0.58
1610	0.59	0.69	1.03	0.89	1.17	1.39
1615	1.12	1.15	1.25	1.41	1.46	1.84
1624	0.95	0.91	0.63	1.03	1.08	1.43
1626	0.63	0.74	0.82	0.61	0.98	1.24
1629	0.39	0.56	0.62	0.69	0.77	0.90
1632	0.68	0.67	0.72	0.76	0.99	1.29
1638	0.84	1.04	0.84	0.75	1.27	1.11
1641	1.09	0.80	0.82	0.95	1.17	1.29
1643	0.48	0.52	0.62	0.72	1.18	1.42
1645	0.81	0.80	0.95	0.84	1.13	1.49
1646	0.86	0.68	0.76	0.82	1.03	1.03
1652	0.58	0.61	0.83	0.99	1.18	1.41
1656	0.67	0.75	0.84	1.00	1.36	1.59
1657	0.46	0.64	0.61	0.78	0.90	1.15
1658	0.42	0.25	0.52	0.63	0.99	0.98
1659	0.83	0.77	0.90	1.01	1.24	1.39
1661	0.53	0.66	0.77	0.88	1.11	1.37
1662	0.62	0.56	0.66	0.75	1.14	1.43
1665	0.31	0.29	0.49	0.70	1.05	1.44
1666	1.30	1.54	1.68	1.63	2.13	2.16
1667	0.66	0.73	0.82	1.21	1.68	2.18
1670	0.72	0.80	1.00	1.12	1.38	1.45
1673	0.64	0.91	1.08	1.08	1.41	1.83
1674	0.65	0.78	0.80	1.03	1.18	1.59
1675	0.52	0.68	0.80	1.06	1.63	1.91

12-3

Sample ID	TIME (hours) 1.5	3	4.5	8	12	24
1535	1.03	1.61	2.61	4.87	7.12	13.09
1536	0.82	1.11	1.62	3.22	4.16	10.55
1538	1.48	1.77	2.39	3.44	5.08	10.99
1539	4.47	8.79	13.58	25.32	33.63	39.18
1545	2.22	4.11	6.90	14.48	19.88	39.71
1564	1.52	2.30	3.49	6.75	11.34	28.48
1572	1.22	1.83	2.92	5.47	8.48	22.03
1587	1.51	2.15	2.96	6.10	9.86	20.96
1590	2.23	3.51	5.81	11.24	16.75	32.54
1592	2.30	4.38	8.52	14.54	22.77	39.64
1512	0.70	1.18	1.83	2.66	3.28	5.17
1519	0.71	0.77	0.97	1.42	2.04	3.30
1521	0.51	0.76	0.99	1.45	2.01	3.89
1524	0.64	0.69	0.82	1.08	1.44	2.24
1527	0.67	0.84	1.06	1.77	2.01	3.77
1530	0.28	0.49	0.57	0.59	0.65	0.76
1548	0.52	0.58	0.80	0.94	1.06	1.40
1550	0.60	0.60	0.58	0.68	0.89	1.10
1589	0.25	0.42	0.38	0.39	0.46	0.61
1606	0.40	0.47	0.42	0.55	0.62	0.91
1610	0.35	0.24	0.33	0.46	0.62	0.68
1615	0.41	0.61	0.61	0.93	0.80	1.05
1624	0.43	0.88	0.56	0.69	0.77	1.01
1626	0.22	0.51	0.59	0.62	0.71	0.96
1629	0.74	0.80	0.84	0.93	1.09	1.37
1632	0.30	0.38	0.40	0.63	0.81	1.04
1638	0.58	0.57	0.60	0.63	0.75	0.93
1641	0.53	0.81	0.81	0.97	1.03	1.53
1643	0.78	1.09	0.87	1.28	1.48	2.27
1645	0.68	0.47	0.46	0.61	0.84	1.05
1646	0.58	0.41	0.64	0.49	0.62	1.00
1652	0.45	0.50	0.61	0.67	0.70	0.90
1656	0.60	0.71	0.78	1.08	1.14	1.71
1657	0.11	0.18	0.29	0.52	0.63	0.76
1658	0.48	0.50	0.64	0.83	1.02	1.26
1659	0.40	0.38	0.59	0.72	0.87	1.34
1661	0.66	0.53	0.65	0.77	0.90	1.37
1662	0.39	0.43	0.65	0.89	1.13	1.77
1665	0.48	0.48	0.63	0.80	1.08	1.65
1666	0.82	0.73	0.71	0.80	1.00	1.56
1667	0.59	0.82	0.84	1.05	1.09	1.76
1670	0.79	0.96	1.00	1.19	1.30	1.77
1673	0.90	1.16	1.09	1.68	1.65	1.93
1674	0.59	0.47	0.56	0.74	0.64	0.95
1675	0.64	0.60	0.83	1.00	1.17	1.64

12-4

Sample ID	TIME (hours) 1.5	3	4.5	8	12	24
1535	0.56	0.75	0.63	0.79	1.04	1.55
1536	0.26	0.29	0.38	0.52	0.67	1.17
1538	0.08	0.07	0.18	0.23	0.32	0.85
1539	0.64	1.09	1.81	2.64	4.28	7.93
1545	0.27	0.61	1.00	1.61	2.28	4.80
1564	0.17	0.41	0.66	1.02	1.80	2.97
1572	0.62	0.73	0.89	1.17	1.72	3.07
1587	0.33	0.41	0.59	0.83	1.38	2.22
1590	0.17	0.45	0.69	1.24	2.15	3.58
1592	0.97	1.27	1.62	2.31	3.34	5.50
1512	1.06	1.09	0.96	1.20	1.20	1.01
1519	1.55	1.89	1.67	1.77	1.81	1.74
1521	0.36	0.19	0.56	0.66	0.91	0.82
1524	0.41	0.48	0.52	0.55	0.51	0.51
1527	0.51	0.62	1.15	0.87	1.07	1.15
1530	0.53	0.47	0.59	0.56	0.76	0.93
1548	0.84	0.74	0.84	0.84	0.99	0.98
1550	0.13	0.29	0.26	0.41	0.37	0.36
1589	0.48	0.46	0.38	0.91	0.64	0.29
1606	0.61	0.49	0.56	0.46	0.42	0.36
1610	1.39	1.55	1.45	1.54	1.85	1.32
1615	0.84	0.84	0.92	0.68	0.94	0.87
1624	0.27	0.34	0.34	0.32	0.10	0.28
1626	0.05	0.42	0.44	0.37	0.37	0.41
1629	0.57	0.54	0.62	0.76	0.83	0.96
1632	0.73	0.66	0.86	0.90	1.77	4.07
1638	0.43	0.35	0.40	0.57	0.59	0.37
1641	0.43	0.38	0.39	0.49	0.50	0.51
1643	0.18	0.26	0.40	0.56	0.65	0.58
1645	1.03	0.82	1.18	0.98	2.42	0.86
1646	0.28	0.38	1.27	0.47	0.48	0.68
1652	0.93	4.54	1.06	1.28	1.04	1.04
1656	0.17	0.31	0.39	0.49	0.55	0.54
1657	0.89	0.90	1.05	0.92	0.92	0.88
1658	0.79	1.16	0.94	1.05	0.78	0.80
1659	1.30	1.09	6.57	1.61	1.46	1.40
1661	0.04	0.03	0.03	0.02	0.27	0.27
1662	0.38	0.33	0.46	0.39	0.44	0.39
1665	0.22	0.50	0.31	0.22	0.28	0.33
1666	0.09	0.12	0.15	0.16	0.19	0.23
1667	0.57	0.54	0.53	0.57	0.65	0.59
1670	0.45	0.58	0.37	0.50	1.30	0.65
1673	0.17	0.17	0.18	0.28	0.40	0.24
1674	0.32	0.33	0.30	0.35	0.46	0.40
1675	0.40	0.46	0.52	0.56	0.59	0.62

Figure 8. Specificity results of top 4 DNAzyme candidates, expressed as percent cleavage values.

At first glance, it is immediately clear that the patient samples used to create the counter selection pool for *in vitro* selection retain very high DNAzyme cleavage percentages, despite aggressive counter selection. Once again, this suggests the possibility of the DNAzyme library misidentifying a common biomarker/target during selection, or could be indicative of a common biomarker with differing concentrations in MHN and MHS serum. Given the pathophysiology of MH, these results more likely depict the former hypothesis, and suggest additional selection rounds with more stringent selection pressures may be required to further tease out biomarker differences between MHN and MHS serum.

Another readily visible trend is the profound difference in cleavage percentage between the *in vitro* selection samples and the new samples obtained from the MHIU, once again suggesting the possibility of overtraining the library to detect the original samples. However, given that this finding is consistent in both the sensitivity and specificity analysis, another explanation for this observation is a true difference in the new serum samples compared to the original samples. The most readily identifiable difference between the original samples and the new samples is in the seroconversion method. The original samples were successfully seroconverted from plasma using calcium, whereas the new samples were seroconverted using addition of thrombin to the plasma. As such, two additional hypotheses emerge that can account for the differences in cleavage observed.

First, seroconversion of the new serum samples may lead to a thrombin-mediated inhibitory effect of the DNAzymes by introduction of a new molecule to the serum samples. This is substantiated by a recent finding of a covalently bound DNA aptamer to thrombin offering the aptamer significant resistance to nuclease degradation.[9] Whether the original serum samples' percent cleavage is target or nuclease-mediated, this hypothesis may account for either molecular interaction being inhibited in the new samples. However, the likelihood of the top four DNAzyme candidates being a molecular match for thrombin is very low, and combined with the low amount of thrombin added to the seroconversion reaction, it is unlikely that the addition of 25 units of thrombin would account for such disproportionate cleavage percentages between the new and original serum samples.

The second hypothesis worth considering to account for the significant differences in DNAzyme cleavage is the presence of EDTA in the plasma samples, as confirmed by the MHIU through the use of EDTA-coated vacutainers for blood sample collection. Since DNAzymes require divalent metal ions to catalyze cleavage reactions, the presence of a chelating agent like EDTA would lead to significant inhibition of DNAzyme function.

The original samples used in the *in vitro* selection protocol were successfully seroconverted with the addition of calcium by presumably oversaturating the EDTA present in plasma until the coagulation cascade was activated. The new samples unfortunately could not be converted using calcium alone, and as a result were seroconverted using thrombin only. Consequently, without extraneous addition of calcium to saturate the EDTA in plasma, it appears the seroconverted samples contain higher concentrations of free EDTA, leading to chelation of calcium and magnesium ions required for DNAzyme catalysis, thus inhibiting DNAzyme cleavage.

While it is difficult to posit why the new samples could not be converted using the same approach as the original selection samples, it is presumably due to their more recent collection. Fresher samples may correlate with a higher concentration of EDTA, and therefore demonstrate a stronger inhibitory effect on the initiation of the coagulation cascade. Since nucleases also rely on divalent metal ions to perform their catalytic function of DNA/RNA degradation,[10,11] the possibility of time-dependent EDTA degradation or a subtle change to the blood collection method is raised as a result of considerable nuclease degradation of DNAzymes observed in samples collected contemporary to the selection samples (1500s) compared to newer samples (1600s), despite seroconversion of both with thrombin. As can be observed in Figure 9, a potentially lower

concentration of EDTA in patient sample 1512 compared to 1658 and 1656 correlates with less chelation of divalent metal ions, and therefore increased nuclease degradation of the 12-1 DNAzyme. The effect of blood collection method and seroconversion on DNAzyme activity and *in vitro* selection will be further explored in Chapter 5.

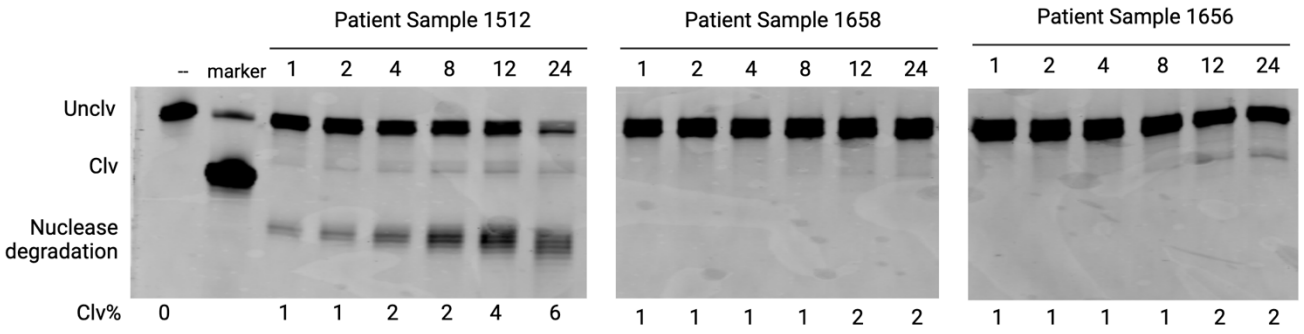


Figure 9. 12-1 MHN sample testing with patient samples 1512, 1658, and 1656. Significant nuclease-driven degradation is observed in 1512 compared to 1658 and 1656. Cleavage percent calculations only reflect the DNAzyme cleavage band at the level of the marker.

In addition to the presence of significant nuclease degradation in patient sample 1512 compared to 1658 and 1656, Figure 9 also demonstrates the skewing effect nuclease degradation has on the percent cleavage calculation. Due to significant and unequal degradation of the uncleaved and cleaved bands over 24 hours, the percent cleavage calculation yields a higher percent cleavage in 1512 despite the cleavage band appearing to have comparable intensity to 1656 and 1658, leading to a false positive result. This effect is observed in multiple MHN samples, which undoubtedly will falsely decrease specificity values.

Returning to the results of the specificity analysis, specificity values were calculated based on the positive test threshold parameters of the sensitivity analysis. Specifically, we previously attributed any percent cleavage above the buffer-only reaction threshold of 1% at 24 hours, to a positive reaction mediated by DNAzyme interactions with serum components. Evidently, 1% DNAzyme

cleavage is a low positive test parameter and unsurprisingly, the specificities of the top three DNAzyme candidates at this positive test threshold are only 18% for 12-1, 11% for 12-2, and 20% for 12-3. Remarkably however, the specificity of 12-4 is considerably higher, suggesting that at the 1% cleavage positive test threshold, 12-4 may be the best candidate to emerge from this analysis with sensitivity and specificity values of 59% and 64%, respectively. The specificity calculation equation[8], along with an exploration of sensitivity and specificity values based on varying positive test cleavage thresholds and incubation time are summarized in Figure 10.

A

$$Specificity = \frac{True\ Negative}{(True\ Negative + False\ Positive)} \times 100$$

B

		24-hour		4-hour	
Positive test threshold	**Sequence**	**Sensitivity**	**Specificity**	**Sensitivity**	**Specificity**
1%	12-1	89%	18%	89%	42%
	12-2	89%	11%	73%	56%
	12-3	95%	20%	68%	69%
	12-4	59%	64%	34%	76%
1.5%	12-1	89%	33%	68%	67%
	12-2	82%	51%	32%	69%
	12-3	73%	44%	39%	76%
	12-4	27%	78%	9%	91%
2%	12-1	70%	56%	52%	71%
	12-2	43%	62%	25%	82%
	12-3	52%	64%	25%	80%
	12-4	18%	82%	2%	98%

Figure 10. A. Specificity[8] calculation equation. **B.** Sensitivity and specificity values for top 4 DNAzyme candidates at varying positive test thresholds and time points. Notable combinations highlighted.

Overall, the sensitivity and specificity analysis has established that each of the top four DNAzyme candidates demonstrates favourable sensitivity and specificity values at specific incubation times and cleavage percentage thresholds. The most notable are 12-3 (sensitivity: 68%, specificity: 69%) and 12-1 (sensitivity: 68%, specificity: 67%), both at 4-hour incubation times. Although none of the candidates meet the gold standard sensitivity and specificity of the CHCT, these preliminary values are very encouraging and suggest that the *in vitro* selection in serum protocol has succeeded. Further optimization of the candidates is warranted for possible inclusions into diagnostic assay development. Such optimization studies include truncation studies for affinity enhancement and determination of functional regions of each DNAzyme with possible reselections to improve sensitivity and specificity.

3.2.4 Kinetics of candidate DNAzyme sequences

A kinetic analysis was performed to further characterize and evaluate the performance of each DNAzyme sequence in each patient sample category. Unlike the pancreatic cancer project, the MHS and MHN sample cleavage time course results have been inputted as replicates in order to generate a generalized rate constant for each candidate DNAzyme in the positive and negative sample categories for ease of comparison. As in the pancreatic cancer project, each reaction was modelled by non-linear regression using the one-phase association equation $Y = Ymax\ [1\text{-}e^{-kt}]$, employing the GraphPad Prism 10.0.3 software. The constraints imposed were $Y_0 = 0$, and Plateau < 100. The emerging rate constants k_{obs} and their associates 95% confidence intervals have been compiled in Table 2.

DNAzyme Sequence	MHS		MHN	
	k_{obs} (hr^{-1})	95% CI	k_{obs} (hr^{-1})	95% CI
12-1	**0.05029**	? to 0.1203	**0.04293**	0.003326 to 0.1194
12-2	**0.04823**	? to 0.1190	**0.05649**	0.002860 to 0.1470
12-3	**0.02063**	? to 0.08222	**0.02786**	? to 0.1066
12-4	**0.3278**	0.2016 to 0.5662	**0.2429**	0.1243 to 0.5229

Table 2. MHS and MHN generalized sample kinetics for top four DNAzyme candidates. ? marks an unknown confidence interval parameter, indicating significant variance in the data.

The data presented in Table 2 aligns with the observations and results discussed in the sensitivity and specificity section. Specifically, 12-1 appears to hold the highest rate constant in MHS serum. This result is in keeping with 12-1 having the highest percent cleavage values for the 24-hour cleavage time courses. When comparing MHS and MHN rate constants, both 12-1 and 12-4 appear to have higher rate constants in MHS compared to MHN, suggesting preferential interaction with MHS serum. This further substantiates the conclusion of *in vitro* selection having successfully yielded sequences capable of distinguishing between positive and counter selection serum samples, and suggests these two DNAzymes hold potential for future incorporation into a diagnostic tool.

3.2.5 Target size parameter identification

Similar to the pancreatic cancer project, we opted to gain more insight into the mediating factors of DNAzyme cleavage in this selection and to narrow in on possible target(s) by performing a size exclusion experiment. This experiment was designed to offer molecular weight parameters for possible targets, and offer any indication whether the MHS DNAzyme targets differ from the

MHN, yielding a potential biomarker of malignant hyperthermia susceptibility. The experiment was performed with pooled MHS and MHN serum using each of the 10 original samples used to conduct the *in vitro* selection. Both categories of serum were centrifuged through filter columns with incrementally decreasing molecular weight cut-offs. The resulting serum fractions were used to perform cleavage time point experiments. Given the robustness of 12-1 demonstrated in this chapter, we felt it would be appropriate to continue exploring it further and employ it in this experiment. The results of this size exclusion experiment are depicted in Figure 11.

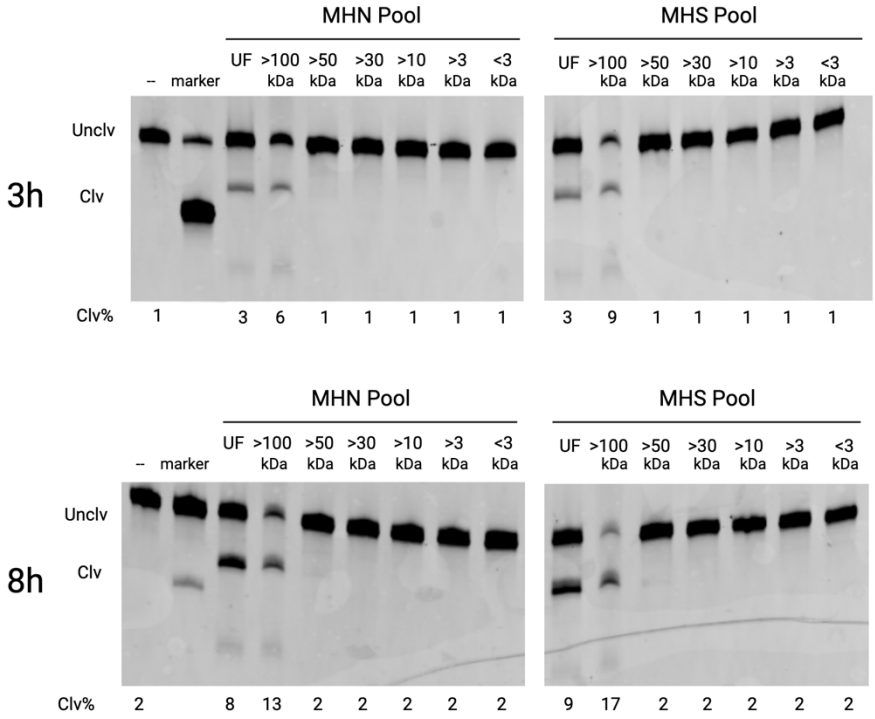


Figure 11. Gel depiction of two time points of the size exclusion fractionated serum cleavage time point experiment.

Interestingly, Figure 11 shows conclusively that cleavage of 12-1 is mediated exclusively by serum components of the largest size fraction, specifically by targets in the >100kDa size range, regardless of positive or counter selection serum category. This finding is confirmed at two

separate time points, with cleavage at 8 hours being predictably higher than at 3h. The finding of both selections' DNAzyme cleavage being mediated by a component in the >100kDa range points to interactions with a common serum component. The most likely mediator of this cleavage is nuclease activity as seen by increased rates of cleavage and degradation in the concentrated >100kDa serum fraction compared to unfractionated serum. However, another possible explanation for the results observed is related to experimental method, and the possibility of the first filter in the queue becoming clogged by the multitude and size of serum components, subsequently malfunctioning and preventing the passage of smaller molecules through. This theory may explain the lack of cleavage or degradation of DNAzyme samples beyond the 100kDa filter through 24 hours, despite the presence of numerous nucleases in serum ranging in size from 10-50kDa.

3.3 Summary

In conclusion, we have successfully isolated four candidate DNAzyme sequences (12-1, 12-2, 12-3, and 12-4) capable of reaching clinically relevant sensitivity and specificity thresholds for detection of MH susceptibility. In doing so, we have once again demonstrated the feasibility of performing *in vitro* selection with DNAzymes directly in whole human serum. While higher values of sensitivity and specificity are observed at 24-hour incubation times with serum samples, two sequences (12-1 and 12-3) show promising trends in sensitivity and specificity beginning at 4 hours, highlighting their potential and merit for further clinical examination. With additional optimization, all four DNAzyme candidates show great promise of reaching clinically significant sensitivity and specificity values, offering a possible alternative diagnostic method for identification of MH susceptibility in the near future.

3.4 Experiments

3.4.1 Enzymes, chemicals, and other reagents

Urea (ultrapure) and 40% polyacrylamide solution (29:1) were acquired from BioShop Canada (Burlington, ON, Canada). The water used was purified via Milli-Q Synthesis A10 water purifier. The enzymes T4 polynucleotide kinase (PNK), T4 DNA ligase, and Taq DNA polymerase were purchased from Thermo Scientific (Ottawa, ON, Canada). Human alpha-Thrombin was purchased from Fischer Scientific (Burlington, ON, Canada). ATP, EvaGreen (20x) and deoxynucleoside 5′-triphosphates (dNTPs) were also purchased from Thermo Scientific (Ottawa, ON, Canada). Silica beads were purchased from Sigma. All other reagents were purchased from Bioshop Canada and used without further purification.

3.4.2 Synthesis and purification of oligonucleotides

A list of the oligonucleotides sequences employed in this selection experiment are listed in Table X. The DNA library MH1, the signalling molecule FQ30, the forward PCR primer MH1-F, the two reverse PCR primers MH1-R and MH1-R2, the template MH1-Splint for ligating FQ30 to MH1 were purchased as synthetic oligonucleotides from Integrated DNA Technologies (IDT). All oligonucleotides were purified by 10% denaturing polyacrylamide gel electrophoresis (dPAGE) prior to use and quantified spectroscopically. Each nucleotide position in the 40 nucleotide random domain of MH1 is randomized by IDT with a 25% probability of A, C, G or T nucleotide. FQ30 contains an adenosine ribonucleotide (rA), flanked by a fluorescein-dT fluorophore and a dabcyl-dT quencher. The reverse primer MH1-R2 contains an 18-atom spacer and a poly-T tail composed of 20 thymine nucleotides at the 5' end. The function of the spacer is to prevent the poly-T tail from being amplified, consequently marking the anti-sense strand with a lengthy poly-T tail to

facilitate recovery of the DNAzyme-coding sense strand. The recovery is accomplished through separation of the two strands by 10% dPAGE. The RNA-containing substrate FQ30 was deprotected and purified by 10% dPAGE following a previously reported protocol.[12]

Selection ID	MH1
Library "MH1-Lib"	5`→3` CGCACCGTAGCAGATGACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCATTGTTTCAGTGGCGAGCA 79nt, 40nt random domain
Forward Primer "MH1-F"	5`→3` CGCACCGTAGCAGATGAC 18nt, Tm~56°C
Reverse Primer "MH1-R"	5`→3` TGCTCGCCACTGAAACAATGA 21nt, Tm~57.4°C
Blocked Reverse Primer "MH1-R2"	5`→3` TTTTTTTTTTTTTTTTTTT/iSP18/TGCTCGCCACTGAAACAATG 40nt, Tm~59.8°C, iSP18 = 18 atom spacer
"FQ30" Substrate	5`→3` CTATGAACTGACXrAYGACCTCACTACCAAG 31nt, X = Dabcyl dT, rA = riboA, Y = Fluorescein dT
"MH1-Splint" Ligation template	5`→3` ATCTGCTACGGTGCGCTTGGTAGTGAGGTC 30nt
i5.int.MH1-F	5`→3` CTTTCCCTACACGACGCTCTTCCGATCTNNNNCGCACCGTAGCAGATGAC 50nt
i7.int.MH1-R	5`→3` GGAGTTCAGACGTGTGCTCTTCCGATCTTGCTCGCCACTGAAACAATGA 49nt

Table 3. Oligonucleotide sequences employed in this experiment.

3.4.3 Plasma sample acquisition and processing

All 90 banked plasma samples were transferred from the Malignant Hyperthermia Investigation Unit (MHIU) at the University of Toronto. Each plasma sample is derived from a single patient. The samples were received in three batches. The first batch contained 20 samples, composed of

10 MH sensitive and 10 MH negative patient samples. Each sample amounted to approximately 1ml of plasma. Upon receipt, samples were seroconverted in 200µl batches, with addition of 2.5µl 1M Calcium Chloride and incubation at room temperature. Samples were checked at 15-minute intervals for clotting. If no clotting was observed, an additional dose of 2.5µl 1M Calcium Chloride was added. Where clotting was not observed at 1h, 10 silica beads were added. Clotting was successfully achieved in all twenty samples in 90 minutes or less. Following clotting, samples were centrifuged at 4°C, 15000 g for 5 minutes. The supernatant (serum) was extracted and aliquoted into microcentrifuge tubes, labelled, and stored at -80°C. Prior to storage, positive and counter selection serum pools were formed and aliquoted. Positive selection serum was formed by mixing equal volumes of each of the 10 MH sensitive patient samples. Counter selection serum was formed by mixing equal parts of each of the 10 MH negative patient samples. The pools were aliquoted, labelled, and stored at -80°C.

The second batch was received two years later, for the sensitivity and specificity analysis. This batch contained 20 samples, composed of 10 MH sensitive and 10 MH negative patient samples. Each sample amounted to approximately 1ml of plasma. Upon receipt, samples were seroconverted in 200µl batches. The first attempt at seroconversion with 2.5µl doses 1M Calcium Chloride and incubation at room temperature was only successful in seroconverting 8/20 samples, despite addition of up to 6-7 doses of 1M Calcium Chloride and addition of silica beads. The successfully seroconverted samples were processed for serum, aliquoted and stored at -80°C. Consequently, a decision was made to repeat seroconversion of all 20 samples in this batch with Thrombin, in order to maintain uniformity in all new testing samples and expedite seroconversion.

All 20 samples were successfully seroconverted with addition of 25U of Thrombin per sample within 30-120s.

The third batch of samples was received shortly following the second batch, and contained 50 samples, composed of 25 MH sensitive and 25 MH negative patient plasma samples. The thrombin seroconversion method was applied, and all 50 samples were successfully seroconverted within 30-120s of addition of 25U of Thrombin.

3.4.4. Fresh collected sample acquisition and processing

One fresh collected sample was acquired through recruitment of a local area patient with a demonstrated history of two separate MH reactions. The recruited patient signed consent forms in accordance with the Hamilton Integrated Research Ethics Board (HIREB). The blood sample was acquired intra-operatively by the writer, using standard blood collection vacutainer tubes and the arterial line in situ. Approximately 10ml of blood was collected in total, into two red-capped vacutainer tubes. Upon collection, the tubes were couriered to the laboratory facility where were allowed to clot at room temperature for 30 minutes. The vacutainers were centrifuged at 4°C, 2200 g for 20 minutes. The supernatant (serum) was aliquoted into microcentrifuge tubes, labelled, and stored at -80°C.

3.4.5. in vitro selection

The *in vitro* selection protocol followed previously described protocols by our research group. Briefly, 800pmol of MH1 was phosphorylated with ATP and 20U of T4 polynucleotide kinase (PNK) in 1× PNK buffer A at 37°C for 20 minutes, in a 100µl reaction volume. The reaction was

stopped by heating the mixture at 90°C for 5 minutes. Water, along with equimolar FQ30 and MH1-Splint were subsequently added to this solution, to a volume of 366µl, followed by heating at 90°C for 1 min and cooled to room temperature to anneal the fragments. Next, 30µl of 10× T4 DNA ligase buffer was added, followed by T4 DNA ligase (20 U) was added, to a total volume of 400µl. The reaction was subsequently incubated at room temperature for 2 h. The DNA molecules in the mixture were concentrated by ethanol precipitation, and the ligated MH1-FQ30 constructs were purified by 10% dPAGE and quantified.

The first round of selection was a positive selection round. Consequently, all 800pmol of the MH1-FQ30 ligated DNAzyme library was mixed with water, heated at 90°C for 3 minutes and cooled at ambient temperature for 5 minutes to allow DNAzyme sequences to fold. 10x Selection Buffer was added to the reaction (500mM HEPES pH 7.0, 1500mM NaCl, 150mM $MgCl_2$) to a final concentration of 1x, followed by addition of the positive selection serum in accordance with the selection conditions detailed in Table 1. The reaction was stopped by addition of 2x Quenching Buffer (2x Urea Loading Buffer + 60mM EDTA), to a final concentration of 1x. The cleaved MH1-FQ30 sequences were purified by 10% dPAGE, and resuspended in 20µl of water. Each selection round gel was scanned using the Amersham Typhoon Biomolecular Imager.

Following positive selection, two PCR reactions (PCR1 and PCR2) were performed to amplify and recover the DNAzyme-coding sense strand. PCR1 was performed with 0.5µM MH1-F and 0.5µM MH1-R primers and 5µl of the purified cleaved MH1-FQ30 from the previous round of positive selection, along with 200 µM each of dNTPs (dATP, dCTP, dGTP and dTTP), 1× PCR buffer (75 mM Tris-HCl, pH 9.0, 2 mM MgCl2, 50 mM KCl, 20 mM (NH4)2SO4) and 2.5 U of

Thermus aquaticus (Taq) DNA polymerase. The DNA was amplified using the following thermocycling steps: 95°C for 1 min; 14-16 (dependent on the amount of cleavage of the PanC2 DNA library) cycles of 95°C for 20s, 52°C for 30s and 72°C for 30s; 72°C for 1 min. For the PCR2 reaction, 5µl of a 1:100 dilution of the PCR1 product was used as the template for this additional PCR step (a total of 48× 50-µL reactions were conducted to generate enough DNA) using primers MH1-F and MH1-R2 and the same protocol as PCR1. The sense strand was purified by 10% dPAGE and used for the next selection round.

Counter selection reactions followed the same protocol as positive selection, with the uncleaved MH1-FQ30 construct instead being purified by dPAGE and carried into the subsequent positive selection round.

3.4.6 Cleavage Time Course Reactions

Cleavage time course reactions with the Counter Selection MHN Serum Pool and Positive Selection MHS Serum Pool using the recovered library from Round 12 were carried out in preparation for sequencing analysis. The reactions were constructed as follows: 144pmol MH1-FQ30 constructs from effectively Round 13 (R13) were incubated with 30%v/v serum, 28.8µl of 10× SB, and water to a reaction volume of 288µl. Time courses were initiated by addition of all components, and incubated at ambient temperature draped with a paper towel for 96 hours. At the indicated time points, a 32µl aliquot was removed from each reaction and mixed with 32µl 2x Quenching Buffer (QB). Time point aliquots were stored at -20°C until completion of the time course. The cleavage was then analyzed by 10% dPAGE, and the cleaved MH1-FQ30 constructs were purified. The image of cleaved and uncleaved DNAzyme bands was obtained with the Amersham Typhoon Biomolecular Imager.

3.4.7 High-throughput sequencing

The cleaved product from the rounds 6-12, along with each individual time point from the R13 cleavage time course was amplified by PCR to obtain sufficient DNA for sequencing. PCR1 was conducted using MH1-F and MH1-R following the same protocol as described above. 5µL of the 1:100 diluted PCR1 product was used as the template for PCR2 using deep sequencing internal primers i5.int.MH1-F/i7.int.MH1-R, and again with assigned external sequencing primers, using the same protocol above for PCR1. 4 individual external primer PCR reactions for each recovered library were performed, and the PCR products were purified by 2% agarose gel electrophoresis. DNA extraction from agarose gel was done using Monarch® DNA Gel Extraction Kit (New England BioLabs). Purified PCR products were sequenced using paired-end Next-Generation Sequencing (NGS) using an Illumina Miseq system at the Farncombe Metagenomics Facility, McMaster University. Raw sequencing reads were first trimmed of their primers using Geneious. The resulting 40 nt reads were filtered for quality using PrinSeq v0.20.4 to make sure only high-quality reads were used for further analysis. 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. 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.

3.4.8 Sensitivity and Specificity Analysis

Following identification of candidate sequences based on enrichment trends identified in the sequencing data, the candidate DNAzyme sequences were ordered from IDT, ligated with the FQ30 segment, and employed in cleavage time course reactions with individual patient samples.

Both the banked samples from the MHIU and the fresh collected sample were tested in the sensitivity and sequencing analysis. The protocol followed the cleavage time course protocol previously described, with 6 time points collected between 1h – 24h. The specific components of the reaction were adjusted as follows: 7pmol FQ30-ligated candidate DNAzyme construct, 30%v/v serum, 1.4µl of 10× SB, and water to a reaction volume of 14µl. Each aliquoted time point was quenched with addition of 2µl 2x QB, and amounted to 1pmol of ligated DNAzyme.

3.4.9 Kinetics Analysis

A kinetic analysis was performed to identify the rate constants of candidate DNAzymes in various patient samples. The data was gathered from the cleavage time courses discussed in the sensitivity and specificity analyses. In accordance with published work examining DNAzyme kinetics (DOI: 10.1002/anie.202012444), each reaction was modelled by non-linear regression using the one-phase association equation $Y = Ymax [1-e^{-kt}]$, employing the GraphPad Prism 10.0.3 software. The constraints imposed were $Y_0 = 0$, and Plateau < 100. A second analysis was performed with the same model and constraints, only with Plateau = 100.

3.4.10 Size Exclusion Experiment

A size exclusion experiment was performed using the most active candidate from the sensitivity and specificity testing (12-1). Two categories of serum were tested, specifically the Counter Selection MHN Serum Pool and Positive Selection MHS Serum Pool. Each pooled serum category was fractionated using size exclusion columns with the following filter size parameters: 100kDa, 50kDa, 30kDa, 10kDa, 3kDa. Beginning with 200µl of each serum pool, serial fractioning was completed by centrifuging the serum with each size column at 4°C, 14000g, for 5 minutes. The

concentrated fraction above the filter was recovered, and the eluent was transferred to the next column in decreasing order of filter size. Cleavage time course reactions were performed by incubating the ligated 12-1 DNAzymes with each recovered serum fraction. The cleavage was then analyzed by 10% dPAGE. The image of cleaved and uncleaved DNAzyme bands was obtained with the Amersham Typhoon Biomolecular Imager.

3.5 References

1. Biesecker LG, Dirksen RT, Girard T, et al. Genomic Screening for Malignant Hyperthermia Susceptibility. *Anesthesiology*. 2020;133(6):1277-1282. doi:10.1097/ALN.0000000000003547
2. Riazi S, Kraeva N, Hopkins PM. Malignant Hyperthermia in the Post-Genomics Era. *Anesthesiology*. 2018;128(1):168-180. doi:10.1097/ALN.0000000000001878
3. Riazi S, Larach MG, Hu C, Wijeysundera D, Massey C, Kraeva N. Malignant hyperthermia in Canada: Characteristics of index anesthetics in 129 malignant hyperthermia susceptible probands. *Anesth Analg*. Published online 2014. doi:10.1213/ANE.0b013e3182937d8b
4. Rosenberg H, Davis M, James D, Pollock N, Stowell K. Malignant hyperthermia. *Orphanet J Rare Dis*. 2007;2(1). doi:10.1186/1750-1172-2-21
5. Litman RS, Griggs SM, Dowling JJ, Riazi S. Malignant Hyperthermia Susceptibility and Related Diseases. *Anesthesiology*. 2018;128(1):159-167. doi:10.1097/ALN.0000000000001877
6. Rosenberg H, Pollock N, Schiemann A, Bulger T, Stowell K. Malignant hyperthermia: a review. *Orphanet J Rare Dis*. 2015;10(1):93. doi:10.1186/s13023-015-0310-1
7. Riazi S, Kraeva N, Hopkins PM. Updated guide for the management of malignant hyperthermia. *Canadian Journal of Anesthesia*. Published online 2018. doi:10.1007/s12630-018-1108-0
8. Shreffler J, Huecker MR. *Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios*.; 2023.

9. Tabuchi Y, Yang J, Taki M. Relative Nuclease Resistance of a DNA Aptamer Covalently Conjugated to a Target Protein. *Int J Mol Sci*. 2022;23(14):7778. doi:10.3390/ijms23147778
10. Dominguez K, Ward WS. A Novel Nuclease Activity that is Activated by Ca $^{2+}$ Chelated to EGTA. *Syst Biol Reprod Med*. 2009;55(5-6):193-199. doi:10.3109/19396360903234052
11. Barra GB, Santa Rita TH, Vasques J de A, Chianca CF, Nery LFA, Costa SSS. EDTA-mediated inhibition of DNases protects circulating cell-free DNA from ex vivo degradation in blood samples. *Clin Biochem*. 2015;48(15):976-981. doi:10.1016/j.clinbiochem.2015.02.014
12. Ali MM, Aguirre SD, Lazim H, Li Y. Fluorogenic DNAzyme probes as bacterial indicators. *Angewandte Chemie - International Edition*. 2011;50(16):3751-3754. doi:10.1002/anie.201100477

Chapter 4. Advanced Bioinformatic Analysis for Evaluation of *in vitro* selection and Identification of Candidate DNAzyme Sequences

4.1 Introduction

The past two chapters have identified two big challenges of the *in vitro* selections performed, which echo the challenges expected of pioneering work. Consequently, both selections were characterized by a lack of assurance as to the true progress of the selection, and lack of guidance/assurance in selecting DNAzyme candidates, given the rather arbitrary default method of selection. As such, we sought to identify whether additional information can be extracted from the multiple Next-Generation Sequencing rounds performed for each selection. We additionally sought to determine whether this analysis can be subsequently methodized into a protocol that can simultaneously inform on the progress of the selection and guide its next steps, including initiation of new selection pressures.

Given the complex and hostile matrix made up by human serum, and the challenge of attempting *in vitro* selection using DNAzymes in serum for the first time, it stands to reason additional tools would be required to essentially diagnosticate and prognosticate the effectiveness of this new endeavour. The bioinformatic analysis to follow was designed to offer insights and answers to two significant questions, namely the efficacy of *in vitro* selection beyond the limited information offered by gel analysis of DNAzyme library cleavage, and whether advanced bioinformatic techniques can offer a more deductive approach to candidate DNAzyme selection, by predicting specificity, cleavage fractions, and even kinetic properties.

4.2 Results

4.2.1 Pancreatic Cancer

4.2.1.1 Analysis design/plan

The pancreatic cancer bioinformatic analysis is based on the results from three separate sequencing runs. The first run followed the 10^{th} round of selection, the second followed the 14^{th} round of selection, and the third followed a cleavage time course reaction with pooled pancreatic cancer serum and the enriched round 14 library, along with separate 4-hour incubations of the round 14 library with each of the counter selection cancer constituents. The first two sequencing runs mainly serve to identify the degree of enrichment across the latter rounds of selection, and to begin identifying prominent candidate DNAzyme sequences through cluster analysis (stratified by identity/structure and function). The third sequencing run serves to evaluate the specificity of any prominent DNAzyme clusters and sequences, point out any particularly selective sequences for pancreatic cancer, and evaluate the feasibility of the selection given the selection pressures applied, in addition to predicting the ability to achieve success – by identifying DNAzyme candidates specific to pancreatic cancer through enrichment pattern analysis.

Notably, this analysis has yielded additional insights not previously expected into other emerging DNAzymes with different scopes, such as generalized cancer detection probes. Through this analysis, we have sought to specifically identify the most enriched sequences at Round 14, identify sequences preferentially selective for pancreatic cancer over other cancers, identify sequences preferentially selective for general cancer biomarkers over normal healthy serum, and estimate kinetic rates of candidate DNAzymes. This was accomplished through four major analyses, namely by comparing the population enrichment across rounds 6-14, comparing population

enrichment across cleavage time points, comparing population enrichment across counter targets, and lastly by estimating kinetic parameters from sequencing data. Multiple dimensions of this data will be explored, particularly the time dimension (cleavage time course and selection of fast-acting and selective sequences for pancreatic cancer), the sequencing and generation dimension (general round-to-round enrichment), and the target dimension (for sensitivity and specificity analysis).

4.2.1.2 Evaluation of DNAzyme library enrichment and selectivity

We begin with an evaluation of the *in vitro* selection efficiency by offering an overview of the evolution of the DNAzyme library through the rounds of selection sequenced, specifically Rounds 6 through 14. By examining enrichment patterns across the latter nine rounds of selection, we can infer whether the selection is responding to the pressures applied and effectively advancing towards the identification of prominent sequences sensitive and specific for the intended target.

Enrichment of the library population is a measure of the ratio of unique sequences within the library examined to the total number of sequences identified.[1] Subsequently, upon receiving sequencing results, the first step of the analysis requires dereplication of the data.[2] Dereplication entails identifying the number and identity of unique sequences in the sequencing read file, along with each sequence's copy number.[2,3] Dereplication enables the calculation of the enrichment ratio of the libraries recovered from every round of selection submitted for sequencing. The equation employed to calculate the enrichment ratio[4] is depicted below and a graphical representation of the enrichment observed over the last 9 rounds of *in vitro* selection is provided in Figure 1.

$$Enrichment\ Ratio = \frac{Total\ number\ of\ sequence\ reads}{Number\ of\ unique\ sequences}$$

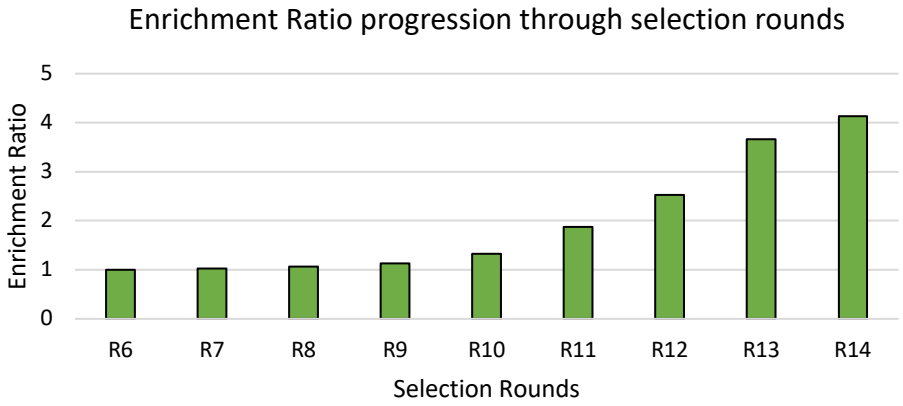


Figure 1. Enrichment Ratio equation and associated depiction of round-to-round enrichment over the course of the *in vitro* selection.

The value of the enrichment ratio indicates the degree to which the selection has progressed. A lower enrichment ratio would indicate slower progression/evolution of the DNAzyme library; conversely, a. high enrichment ratio would indicate rapid progression of the library through the selection protocol and provide confirmation that the selection pressures are effectively paring down the library towards enrichment of functional sequences. Figure 1 therefore confirms steady enrichment of the recovered libraries through the last 9 rounds of selection in favour of decreasing total numbers of sequence reads and unique sequences. Consequently, the expectation for *in vitro* selection is the gradual emergence of functional sequences with increasing copy numbers, and related increase in the fractional percentage of those sequences within the total library population. The fractional percentage of a sequence within a population will further be referred to as the frequency of a sequence within a library population. The equation employed to calculate the frequency of a sequence is depicted below.

$$Unique\ Sequence\ Frequency = \frac{Copy\ number\ of\ unique\ sequence}{Total\ number\ of\ sequence\ reads}$$

By comparing frequencies of dereplicated sequences across multiple rounds of selection, we can track enrichment of a DNA library over time. Log-Log sequence frequency plots are an effective method of comparing sequence frequencies between two conditions, such as two different rounds of selection. In a log-log frequency plot, two DNA libraries' dereplicated sequences are compared against each other, with one library plotted on the x-axis, and one on the y-axis. Each dereplicated sequence within each DNA library is assigned a data point with (x,y) coordinate values which correspond to the sequence's frequency in each of the two libraries being compared. The axis' scales are logarithmic, such that the intersection represents 10^0 (100% abundance) in each library, and as the axis diverge, the frequency values decrease. Consequently, the most enriched (and therefore most significant) sequences will gravitate towards the bottom left of the plot, given their higher frequency of detection. Conversely, most data points will spread towards the top right of the plot due to fewer detections of each individual sequence and more sequence variability. In addition, a diagonal line (y = x) is plotted to facilitate data interpretation. Sequences falling on the diagonal represent no change in frequency in the DNA libraries compared. Sequences plotted below the diagonal demonstrate enrichment in the y-axis library, whereas sequences above the diagonal demonstrate higher frequency in the x-axis library. Figure 2 depicts two data models for detecting enrichment in the PanC2 library, by comparing dereplicated sequence frequencies over multiple rounds of selection against a common baseline (Round 6 library), and through round-to-round sequence frequency comparisons.

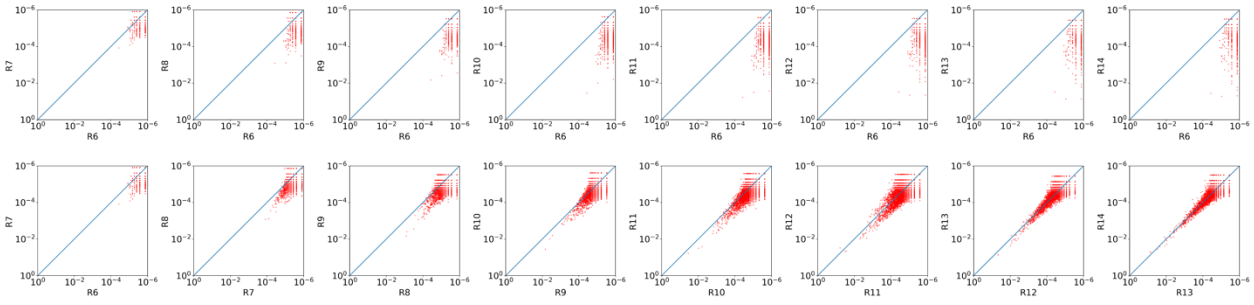


Figure 2. PanC2 Selection – Sequencing Data – Population Progression Log-Log sequence Frequency Plots. **A.** Top Row - Comparison of Rounds 7-14 populations to Round 6. Populations are displaced below the diagonal in each comparison indicating enrichment of sequences. **B.** Bottom Row – Comparison of sequential rounds R7 v. R6 to R14 v. R13. Populations are displaced below the diagonal on all comparisons, except R14 v. R13.

Both models in Figure 2 demonstrate the same conclusion – namely that enrichment of the DNAzyme library appears to have been most rapid in earlier rounds of selection, followed by a tapering in later rounds. This suggests that by Round 13, the library is becoming less responsive to the selection conditions applied, and consequently the selection pressures may need to be adjusted. Specifically, the top row of Figure 2 shows a steady downward migration of the dereplicated sequences over the course of the selection with a tapering in the migration noted at Round 13, suggesting the library was appropriately responding to the selection pressures applied throughout the selection protocol and appropriately enriching.

The bottom row of Figure 2 offers more insights into the round-by-round effects of the selection pressures applied by comparing consecutive pairs of selection rounds. These comparisons, when correlated with the selection conditions applied at each round, can infer what experimental condition gives us what amount of change in the population, for the purpose of optimizing the selection conditions of the experiment. Specifically, when observing the data from the Round 6 and 7 comparison, the bulk of the sequences are located below diagonal, indicating that enrichment

is occurring between Round 6 and 7. This enrichment trend is seen until the Round 11 and 12 comparison, whereby the bulk of the samples appear to align more with the diagonal. However, it is worth noting that the most abundant sequences in Round 12 (bottom left) are still falling below the diagonal. This observation indicates that while enrichment is slowing down compared to previous rounds, there is still enrichment occurring at this round. However, by Rounds 13 and 14, the distribution of the sequence coordinates largely follows the diagonal line, indicating that the library population has stabilized at the selection conditions and pressures applied. Interestingly, beginning at Round 10 more aggressive counter selection measures were implemented with a peak counter selection incubation time of 48 hours. The enrichment plateau therefore harsh newly implemented selection pressures therefore may indicate exhaustion of the library and a natural stopping point for the selection.

Following analysis of enrichment across the latter nine rounds of selection, we turned our attention to another comparison, specifically enrichment of the Round 14 library across multiple time points, in an effort to identify any fast-acting sequences. Figure 3 depicts the results of this analysis. Once again, individual dereplicated sequences have been assigned corresponding frequencies based on their number of detections in each time point, and Log-Log sequence frequency plots have been employed to display the data generated. Similar to Figure 2, the top row represents multiple time point distributions across a common baseline (96h incubation time point), while the bottom row depicts time point-by-time point comparisons.

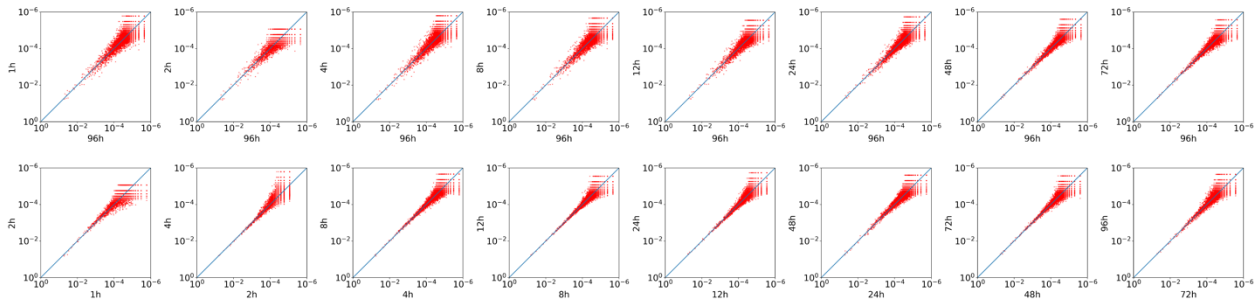


Figure 3. PanC2 Selection – Sequencing Data – Timepoint Comparison. Log-Log sequence Frequency Plots of Round 14 Population. **A.** Top Row - Comparison of timepoints 1-72h vs. 96h. **B.** Bottom Row – Comparison of sequential timepoints 1h v. 2h to 72h v. 96h.

The first observation that can be made from the top row is the degree of data spread in the early time points compared to the 96h time point, as opposed to a more compact distribution in the later time points. This variation in the data spread across time points is attributed to a wider variation of data when comparing early time points to the terminal time point. This effect amounts to larger cumulative changes observed in the 1h and 96h time point comparison as opposed to the 72h and 96h time point comparison, since the majority of the DNAzyme sequences' cleavage has occurred by 24-48h. This effect is substantiated in the bottom row, where most consecutive time point comparisons demonstrate compact data distributions, suggesting minimal changes across consecutive time points and stabilization of the populations. The purpose of comparing widespread time points, as previously alluded to, serves to identify any sequences or classes of sequences that show faster reaction rates, or conversely slower rates. When specifically comparing the 1h vs. 96h time points, the three most frequent sequences (bottom left) appear to fall slightly below the diagonal, suggesting a faster rate of reaction.

In addition to highlighting faster-acting sequences, this time point analysis can help form a basis for clustering sequences for more in-depth analysis towards identification of promising candidate

DNAzymes. Although the process of clustering can vary based on which attribute is most valued in the analysis (identity, function, rate), a balance of identity and function seem to represent the optimal criteria. This may manifest as the presence of a functional motif correlating to faster activity, such that sequences sharing this motif will share similar functionality both in terms of rate of reaction and potential target compatibility, without necessarily sharing 90% identity (as is the default for identity-based clustering). Ultimately, functional need is the most important attribute for driving clustering, in order to effectively reveal distinctive sensitivity and specificity trends. Relating back to the time-point comparison data, the three most frequent sequences in the 1h vs. 96h comparison plot may constitute different clusters, such that we expect most of the sequences belonging to each cluster to similarly fall below the diagonal and share similar rates of reaction.

Notable differences in sequence frequency across sequential time points (bottom row of plots of Figure 3) appear to be limited to the first time point comparison – specifically 1h vs 2h. The same three most common sequences highlighted in the 1h vs 96h plot appear to fall below the diagonal in this first sequential time point comparison as well, suggesting their reaction rate is fast enough to distinguish between the one and two hour time point.

Having explored the round-to-round generation dimension and time dimension of the sequencing data obtained, we next turned our attention to the target dimension. Similar to the two previous dimension analyses, this analysis was also comprised sequence frequency comparisons across multiple different targets, namely pools of pancreatic cancer, breast cancer, colorectal cancer (CRC), lung cancer, prostate cancer, normal human serum (NH), and fetal bovine serum (FBS) samples. The library employed for these comparisons is the enriched Round 14 library, as per the

time course analysis above. The data for the pancreatic cancer sequence frequencies was therefore obtained from the time course data, by extracting the 4h time point to match the incubation time with the other targets. The results of this analysis are illustrated in Figure 4. Once again, the data was plotted on Log-Log Frequency Plots, where the top row is a comparison of each target against pancreatic cancer, and the bottom row is a comparison of each target against normal human serum.

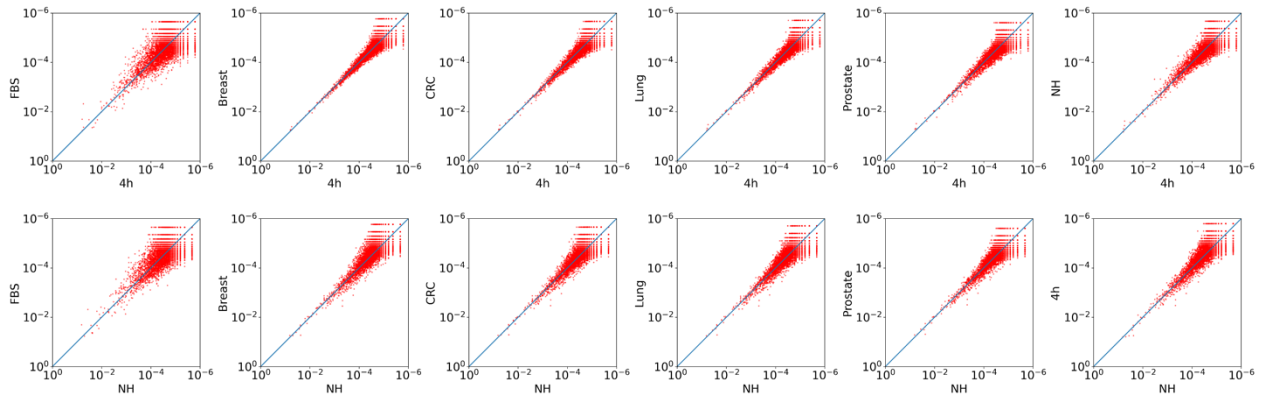


Figure 4. PanC2 Selection – Sequencing Data – Target Comparison. Log-Log sequence Frequency Plots of Round 14 Population, 4h timepoints with counter selection targets. **A.** Top Row - Comparison of FBS, Cancer Serums and Normal Serum to 4h (Pancreatic Cancer) reference population. **B.** Bottom Row – Comparison of FBS and Cancer Serums to Normal Serum.

At first glance, notable observations can be made regarding data spread across different targets, as well as outliers in terms of sequence frequencies. FBS comparisons appear to demonstrate the most widespread data points, consistently so in both pancreatic cancer and normal human serum comparisons, suggesting FBS is the target responsible for the data spread. This is of course consistent with interspecies serum variation, particularly in the composition and ratios of serum nucleases. As compared to FBS, the other human targets (cancer and healthy) are generally more compact along the diagonal, with some minor biases towards specific targets both in the bulk of the data points and in the highest frequency sequences. For instance, the prostate cancer vs. pancreatic cancer comparison demonstrates a subtle downshift of the data points below the

diagonal, suggesting mild enrichment towards prostate cancer. Similar patterns can be seen with lung and colorectal cancer, although the differences appear to be quite subtle, particularly when compared to normal human serum, which shows more data spread.

The overall impression from the pancreatic cancer baseline comparisons suggests similar library population behaviour across all cancer pools compared to normal serum. This observation is inherently expected, matching the sensitivity and specificity analysis results, which demonstrated poor specificity of the candidate DNAzyme probes for pancreatic cancer against other cancer types. This conclusion is further substantiated by the data observed in the bottom row of log-log sequence frequency plots in Figure 4, where we note comparatively more widespread data across all cancers when compared to normal, indicating that the cancer populations all behave similarly to each other. One important caveat worth noting is the lack of multiple normal human samples, making it difficult to conclude with certainty whether the Round 14 DNA library population behaviour exhibited in the comparisons with normal human serum are representative of the general healthy patient population, or possibly specific to the one sample employed in this incubation experiment.

Another notable observation from Figure 4 is the presence of 1-3 high frequency sequences consistently present below the diagonal in multiple cancer comparisons to normal human serum. These sequences warrant further exploration for possible specificity for one cancer type or perhaps as general cancer biomarker detection probes. These sequences and their associated clusters will be discussed further in the next section.

Overall, by identifying individual sequence frequencies, we were successfully able to track the progress of the *in vitro* selection protocol through multiple dimensions – round to round, over time, and across multiple targets. Each analysis so far demonstrated proof of enrichment, indicating that applying selective pressures and an *in vitro* selection protocol to a DNA library in whole human serum is a feasible endeavour.

4.2.1.3 Evaluation of sequence clusters for identification and projection of promising candidate DNAzyme sequences

Following enrichment analysis, we proceeded to cluster the last selection round's DNA library (Round 14 population), in preparation for cluster analysis towards identification of auspicious candidate DNAzyme sequences. In the absence of any indication of a functional motif, or distinctive functionality across different targets, the sequences were clustered by their identity, with a threshold of 90% homology, or within 4 nucleotide mutations of each other. The top 50 sequences ordered by frequency at Round 15 are displayed in Figure 5, along with their associated cluster rank.

Despite an enrichment plateau previously observed in Rounds 13 and 14, Figure 5 denotes high overall population diversity persisting through Round 14, with the top sequence holding a frequency value of only 5.3% (Cluster 3). Additionally worth noting is the finding of most sequences in Figure 5 still enriching, as evidenced by their frequency values peaking in Round 14, with the exception of the top 3 sequences (and clusters), all of which appear to be in decline. Moving forward, given the high diversity of the population, stratifying the population by cluster may prove more effective in detecting groups of related sequences with a similar behaviour trend.

By identifying multiple related sequences with similar functionality, we can effectively increase the confidence in the observed effect. Consequently, cluster analysis is best suited to extract additional trends within the Round 14 population, and identify promising candidate DNAzyme sequences with high sensitivity and specificity for pancreatic cancer.

Sequence	Cluster Rank	R6	R7	R8	R9	R10	R11	R12	R13	R14	Trend
CCATGCACGGTTTTGGACAAATAAGTGGGGTTGTTAGTCG	3	0.0000	0.0001	0.0008	0.0039	0.0099	0.0321	0.0459	0.0737	0.0527	
CCTGGTGTCGACTAGTTCACTTGAGGTTGTTAGTCGGTAT	2	0.0000	0.0000	0.0003	0.0028	0.0063	0.0267	0.0454	0.0728	0.0497	
CAGCCATTGGTTGTATGGAGGTTCGATGTAAACTGTGAGGG	1	0.0000	0.0001	0.0008	0.0068	0.0355	0.0602	0.0304	0.0300	0.0365	
TGTGGCAGTCAGAAACTCGAGGTTGTTAGTCGGGTCCTGA	5	0.0000	0.0000	0.0001	0.0007	0.0016	0.0049	0.0096	0.0227	0.0258	
AGCAGTATCATCTAGTTAACATGAGTTTTAGTCAGACGCC	7	0.0000	0.0000	0.0000	0.0003	0.0005	0.0026	0.0048	0.0120	0.0240	
CAATCACCGTTCAAAGCAACTATTATGGTATGGTTGTTAG	4	0.0000	0.0000	0.0000	0.0001	0.0003	0.0015	0.0037	0.0094	0.0189	
CATGCCTAGTTGTTGGGTAGAGTAACGTGGAGGTTGTTAG	8	0.0000	0.0000	0.0000	0.0003	0.0008	0.0033	0.0082	0.0165	0.0167	
CGCACATCCTCAGTATCTTAGAGGTTGTTAGTCTGGACAA	9	0.0000	0.0000	0.0001	0.0003	0.0008	0.0030	0.0052	0.0095	0.0123	
ACTCCGATTGGGCGATTACGGCCATAAGTACGGTTCTTAG	10	0.0000	0.0000	0.0001	0.0005	0.0008	0.0031	0.0066	0.0103	0.0108	
ACTGTGCTGTCATCGGTGTATAAACTTCGGAGGTTGTTAG	12	0.0000	0.0000	0.0001	0.0004	0.0008	0.0032	0.0058	0.0110	0.0091	
CCATGCAATCGTTAACGCTGAAGATGTAGTGTGTTTTTAG	14	0.0000	0.0000	0.0000	0.0001	0.0003	0.0013	0.0027	0.0053	0.0083	
CACGGCTGCAAAATTCTCTCGTTAACTGACATAATCCGTC	6	0.0000	0.0000	0.0002	0.0009	0.0014	0.0021	0.0088	0.0064	0.0080	
CCTGGTGTCGGCTAGTTCACTTGAGGTTGTTAGTCGGTAT	2	0.0000	0.0000	0.0001	0.0004	0.0010	0.0046	0.0082	0.0126	0.0078	
CCCACCACGTCAGATCAATCATAACGCCTTACTGATGGAC	11	0.0000	0.0000	0.0000	0.0001	0.0003	0.0006	0.0029	0.0039	0.0076	
CTAGCTAGGGATTCTTGCAAAGTACGGTTCCTTAGTTGGT	13	0.0000	0.0000	0.0001	0.0004	0.0006	0.0023	0.0047	0.0061	0.0068	
CGAGGTTGAGTTTGTCTGCCTTAAGATGTACGGTTTTTAG	17	0.0000	0.0000	0.0000	0.0001	0.0003	0.0011	0.0023	0.0039	0.0046	
CATCGGAATCAGTTCTAGCTACGTCGTCATCGTGAAAAGG	33	0.0000	0.0000	0.0000	0.0001	0.0001	0.0006	0.0018	0.0037	0.0043	
CCGCGTAGTACTCTCGCAATAATTCGTCAACTGATCCACG	18	0.0000	0.0000	0.0001	0.0004	0.0005	0.0010	0.0035	0.0036	0.0040	
CCATGCACGGTTTTGGACAAATAAGTGGGGTTGTTAGTCA	3	0.0000	0.0000	0.0000	0.0001	0.0003	0.0014	0.0023	0.0040	0.0038	
CGTCATGGAGATACAGTAGTTTGGGTGGTTGTTAGTTTGT	15	0.0000	0.0000	0.0001	0.0008	0.0015	0.0035	0.0053	0.0061	0.0038	
CGCCATGTCATTCATCCTCTCAGCCTTTGAAATTGCCAGA	16	0.0000	0.0000	0.0000	0.0001	0.0001	0.0004	0.0017	0.0027	0.0037	
ACGTAGTTCGAGTTGTCTCATCGGTCTTGGAGGTTGTTAG	20	0.0000	0.0000	0.0000	0.0001	0.0003	0.0008	0.0019	0.0037	0.0035	
GCACATTACCGTTTGACTTGGAGGTTTTTAGTCGTAATCC	22	0.0000	0.0000	0.0000	0.0002	0.0005	0.0014	0.0022	0.0037	0.0031	
GCATGATGTAAGTAACTGGCATTAGGAGTGATGTTGTTAG	31	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0011	0.0028	0.0031	
CGGCAGGTTAATGGGATAGCAGTCTTGGTATGGTTGTTAG	27	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0.0020	0.0027	0.0028	
ATCACCCCAAACGTCTGTTATCGTAGCCGGTTGTTAGTCT	23	0.0000	0.0000	0.0001	0.0004	0.0007	0.0022	0.0031	0.0035	0.0026	
CTGTCCTATCCGGAAATAGTATGTCTTGGGGGGTTGTTAG	47	0.0000	0.0000	0.0000	0.0001	0.0002	0.0008	0.0016	0.0024	0.0025	
ACACAATGTGCAAGTAACTCGCGAAGTAGAAGGTTGTTAG	28	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0008	0.0015	0.0024	
CCATCGTCATCATGAAATAAACGGGTAGGCAACTTCATCC	39	0.0000	0.0000	0.0000	0.0001	0.0001	0.0007	0.0016	0.0019	0.0024	
CAAGTCACGGAAAGTATCGTTCGTTTGGTCAGGTTGTTAG	43	0.0000	0.0000	0.0000	0.0001	0.0003	0.0006	0.0013	0.0024	0.0024	
TCACTCCAACTGTAATTTGAACTGTTAGTATGGTTGTTAG	49	0.0000	0.0000	0.0000	0.0001	0.0002	0.0005	0.0009	0.0015	0.0022	
CAACCACGTAGGGATTGTTAGTTTAAATTTAGCCAAGTGA	68	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0007	0.0010	0.0022	
ATACCAGATCGTTAGTGTGCAGTCACTTGGAGGTTTTTAG	37	0.0000	0.0000	0.0001	0.0004	0.0005	0.0015	0.0020	0.0030	0.0021	
CATGCGCAGTTCAATAAATCTTGCCTTTGAACTAGTTACC	73	0.0000	0.0000	0.0000	0.0001	0.0001	0.0004	0.0011	0.0018	0.0020	
CAGGCACAGGTTTGTTCTATTCGTTGACTGACAACGGCA	26	0.0000	0.0000	0.0001	0.0002	0.0002	0.0005	0.0017	0.0017	0.0019	
ACAAGGGCATACGTCAGATAGTTTTTATCGTTACTGTCAC	32	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0010	0.0019	
ACGCGGACAATCTGTCTTAGTTAGTTCAAGTAGTGCGTGA	59	0.0000	0.0001	0.0003	0.0005	0.0007	0.0012	0.0012	0.0015	0.0019	
CCCGGTGTCGACTAGTTCACTTGAGGTTGTTAGTCGGTAT	2	0.0000	0.0000	0.0000	0.0001	0.0003	0.0012	0.0019	0.0028	0.0018	
GGGTAGAACAATTCACTTCAGTTTAACTCGAGGTTGTTAG	25	0.0000	0.0000	0.0000	0.0001	0.0001	0.0007	0.0011	0.0015	0.0018	
ACCACGTCAGATATCCATTCGTTCCTCTGAAAGTTCATGG	44	0.0000	0.0000	0.0000	0.0001	0.0001	0.0003	0.0008	0.0014	0.0017	
ACCAGGCATGAATAAGTTGGAGATTATGTTTGGTTGTTAG	30	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0011	0.0016	0.0017	
CACGAACGGAATGCCGGATATTTAGCTGTGTGGTTGTTAG	69	0.0000	0.0000	0.0000	0.0001	0.0002	0.0006	0.0010	0.0012	0.0016	
TAGTCGAAGGGAATTACGTCACATGAATATTATGCCATCC	54	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0006	0.0011	0.0016	
CCAGCCTTGATAGTCAGATTCTAAGCTTATTGAAATGGGC	65	0.0000	0.0000	0.0000	0.0001	0.0001	0.0002	0.0006	0.0008	0.0016	
CCCGGTGAAGATAGGAACATGCAGGTTGTTAGTCATTCGC	67	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0005	0.0011	0.0015	
TGTTGACATAGAGAGAACTGTTAAGGGTTGCGGTTGTTAG	52	0.0000	0.0000	0.0000	0.0001	0.0001	0.0003	0.0006	0.0012	0.0015	
TCGGCCTACATCTTTGATATGAACGGAGGTTCTTAGTGGT	41	0.0000	0.0000	0.0001	0.0003	0.0007	0.0014	0.0017	0.0021	0.0015	
CTGCTAATTCGGAACTGACCTGGGGGAGTATGGTTGTTAG	64	0.0000	0.0000	0.0000	0.0001	0.0001	0.0005	0.0008	0.0013	0.0015	
CTCCTCGTCAAGTGAAATTAAAGGACAGTACAGTTTCCGC	48	0.0000	0.0000	0.0000	0.0001	0.0001	0.0003	0.0010	0.0012	0.0015	
CGGAGAATGTGAATCACCTCCAGTTCAACAGTCCATTGAA	127	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0010	0.0014	

Figure 5. PanC2 Selection – Sequencing Data – Round 14 Ranking. Top 50 sequences in Round 14 with associated cluster ranks.

Following cluster stratification of the Round 14 library population, each individual cluster's performance was evaluated across multiple targets to identify clusters or sequences within clusters specific for a particular target. For ease of concept illustration, the cluster analysis to follow will compare cluster behaviour in colorectal cancer and normal human serum, given a greater observed disparity in the activity of the clusters between these two targets. Similar to the previous subsection, cluster performance is best visualized through Log-Log Sequence Frequency Plots. To contextualize the behaviour of the cluster and identify specific trends in selectivity and specificity, the data points of the clusters profiled in each target comparison will be layered against the performance of the Round 14 population. Figure 6 illustrates the performance of 16 clusters in colorectal cancer versus normal human serum.

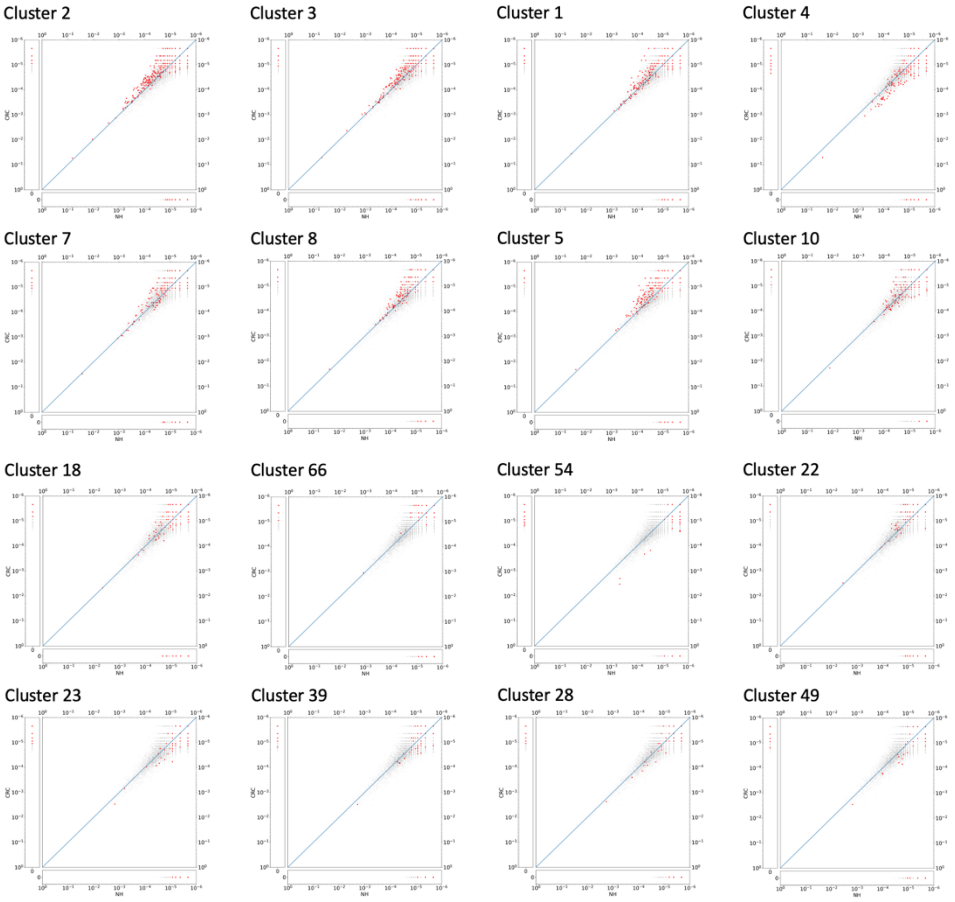


Figure 6. PanC2 Selection – Sequencing Data – CRC v. Normal Serum. Log-Log Sequence Frequency Plot. Grey points: Round 14 Population. Highlight By Cluster – Red points.

Beginning with Cluster 2, it is readily apparent that the majority of the sequences skew above the diagonal line, suggesting most of the cluster is enriched in normal human serum and depleted in colorectal cancer serum. Consequently, we can conclude that Cluster 2 is not specific for colorectal cancer detection. Similarly, Cluster 3 sequences appear to spread above the diagonal. Yet, a high-frequency sequence can be noted at the $\sim 10^{-1}$ frequency in both serums, suggesting it to be the cluster representative, based on it holding the highest frequency values. However, given its position on the diagonal line, its specificity for colorectal cancer is presumed to be low.

Conversely, the Cluster 4 representative sequence is noticeably below the diagonal, with the rest of the cluster following the same pattern of enrichment but at lower frequencies. Therefore, we can conclude that Cluster 4 shows specificity towards colorectal cancer when compared to normal human serum. Clusters 7, 8, and 5 show similar patterns to Cluster 2, and therefore suggest slow specificity and enrichment in colorectal cancer serum, while the representative sequence of Cluster 10 shows some specificity for colorectal cancer despite the rest of the cluster spreading widely across the diagonal at lower frequencies. One final cluster worth highlighting is Cluster 54, whereby the majority of the sequences composing the cluster fall significantly below the diagonal, suggesting the majority of the sequences in this cluster show specificity towards colorectal cancer detection. Lastly, as we progress through individual cluster analysis, it is worth noting that as the cluster rank numbers increase, the number of sequences within each cluster eventually decrease. This effect is attributed to a decreasing number of total sequences left over with each cluster formed by the clustering algorithm.

A readily apparent challenge of analysing clusters of sequences through log-log sequence frequency plots is the difficulty in comparing cluster performances to each other and across multiple targets in quantifiable ways (without sifting through numerous plots). Therefore, we generated a standardization calculation – the mean cluster enrichment ratio, as a quantifiable measure to facilitate ranking of clusters based on the intended specificity targeted. In order to calculate the mean cluster enrichment ratio, we first identified the mean (or average) frequency of the sequences in a particular cluster. The mean frequency of the cluster is subsequently normalized against different targets of interest (baselines), generating mean cluster enrichment ratios, such that values above 1 correlate to clusters falling below the diagonal of a log-log plot indicating enrichment in the target of interest. Conversely, ratios below 1 would indicate a lack of enrichment in a particular target, reflecting a log-log sequency frequency plot position above the diagonal line of no change.

Subsequently, this normalization strategy allows for easier comparison among clusters, as well as tracking their performance across multiple different targets. Figure 7 depicts a standardized summary of the performance of the top 100 clusters across the multiple serum categories employed in this selection. The mean frequency of each cluster (cumulative cluster frequency) was calculated for each cancer type making up the counter selection pool, as well as the pancreatic cancer pool (marked R14 in Figure 7). Next, the mean frequency of each cluster was normalized against the mean frequency in normal human serum, followed by a normalization of pancreatic cancer against the pool of counter selection cancers to create the mean cluster enrichment ratios. These two normalizations serve to identify clusters showing specificity to any kind of cancer over normal human serum, as well as specificity for pancreatic cancer over other cancer types. Consequently,

based on the value of the mean enrichment ratio (above or below 1) as well as the magnitude of the ratio, we can infer the degree of specificity between the targets compared.

Cluster Parent Sequence	Cluster Rank	Cumulative Cluster Frequency					Mean Cluster Enrichment Ratio						Tested (R14 Rank)
		CRC	Breast	Lung	Prostate	R14	CRC/NH	Breast/NH	Lung/NH	Prostate/NH	R14/NH	R14/Counter	
CAGCCATTGGTTGTATGGAGGTTCGATGTAAACTGTGAGGG	1	0.0457	0.0514	0.0431	0.0396	0.0469	1.0723	1.0476	1.0020	1.0251	1.2857	1.2243	
CCTGGTGTCGACTAGTTCACTTGAGGTTGTTAGTCGGTAT	2	0.0801	0.0804	0.0834	0.0789	0.0718	0.9350	0.8167	0.9093	1.0366	1.0576	1.3856	
CCATGCACGGTTTTGGACAAATAAGTGGGGTTGTTAGTCG	3	0.0679	0.0718	0.0636	0.0640	0.0678	0.8887	0.8716	0.8699	0.9669	1.0745	1.2786	
CAATCACCGTTCAAAGCAACTATTATGGTATGGTTGTTAG	4	0.0610	0.0633	0.0615	0.0583	0.0224	1.6643	1.5015	1.5422	1.6419	0.9521	0.7444	6
TGTGGCAGTCAGAAACTCGAGGTTGTTAGTCGGGTCCTGA	5	0.0240	0.0251	0.0247	0.0304	0.0308	0.8167	0.7065	0.8125	0.9864	1.1768	1.5911	4
CACGGCTGCAAAATTCTCTCGTTAACTGACATAATCCGTC	6	0.0182	0.0148	0.0170	0.0223	0.0105	0.9255	0.6974	0.8824	1.1165	0.9238	1.0654	
AGCAGTATCATCTAGTTAACATGAGTTTTAGTCAGACGCC	7	0.0369	0.0377	0.0366	0.0409	0.0299	1.0154	0.9770	1.0312	1.2056	1.1306	1.1592	5
CATGCCTAGTTGTTGGGTAGAGTAACGTGGAGGTTGTTAG	8	0.0250	0.0273	0.0257	0.0219	0.0197	0.8666	0.8053	0.8761	0.8332	0.9012	1.3273	7
CGCACATCCTCAGTATCTTAGAGGTTGTTAGTCTGGACAA	9	0.0173	0.0186	0.0170	0.0184	0.0155	0.8812	0.8225	0.8224	1.0501	0.8927	1.2686	8
ACTCCGATTGGGCGATTACGGCCATAAGTACGGTTCTTAG	10	0.0215	0.0202	0.0192	0.0200	0.0124	1.3297	1.0366	1.1249	1.2068	1.1153	1.0090	9
CCCACCACGTCAGATCAATCATAACGCCTTACTGATGGAC	11	0.0096	0.0098	0.0103	0.0093	0.0095	1.3778	1.1966	1.2750	1.3906	1.5399	1.4890	14
ACTGTGCTGTCATCGGTGTATAAACTTCGGAGGTTGTTAG	12	0.0126	0.0126	0.0127	0.0119	0.0109	0.8762	0.8274	0.9128	1.0547	1.0038	1.3058	
CTAGCTAGGGATTCTTGCAAAGTACGGTTCCTTAGTTGGT	13	0.0126	0.0111	0.0114	0.0121	0.0082	0.8837	0.8316	0.8280	1.0174	0.9758	1.1477	15
CCATGCAATCGTTAACGCTGAAGATGTAGTGTGTTTTTAG	14	0.0112	0.0110	0.0105	0.0109	0.0100	1.0560	0.8357	1.0131	1.0170	1.1674	1.2601	11
CGTCATGGAGATACAGTAGTTTGGGTGGTTGTTAGTTTGT	15	0.0064	0.0056	0.0062	0.0062	0.0049	0.9624	0.8211	0.9079	0.9249	1.0494	1.4226	
CGCCATGTCATTCATCCTCTCAGCCTTTGAAATTGCCAGA	16	0.0053	0.0055	0.0062	0.0048	0.0047	1.3216	1.2959	1.5001	1.3954	1.4586	1.3213	
CGAGGTTGAGTTTGTCTGCCTTAAGATGTACGGTTTTTAG	17	0.0093	0.0082	0.0089	0.0090	0.0056	1.0461	0.8514	0.9601	1.1287	1.0585	1.1668	16
CCGCGTAGTACTCTCGCAATAATTCGTCAACTGATCCACG	18	0.0062	0.0054	0.0057	0.0067	0.0051	1.1286	0.9217	1.0198	1.1937	1.0047	1.2615	
CCCACGATTTCGTCAACTGAAACGTCAATTTCTCATCCCG	19	0.0031	0.0027	0.0029	0.0034	0.0019	1.5491	1.2050	1.1922	1.6470	1.4548	1.1561	
ACGTAGTTCGAGTTGTCTCATCGGTCTTGGAGGTTGTTAG	20	0.0052	0.0053	0.0049	0.0043	0.0044	0.8082	0.7081	0.7898	0.8931	0.8716	1.2177	
CGCGTAGCTAGAATTCTCGTCGACTGAATATTCCCGTTCC	21	0.0025	0.0020	0.0024	0.0027	0.0014	1.1166	0.9745	1.1465	1.3326	0.9459	0.9288	
GCACATTACCGTTTGACTTGGAGGTTTTTAGTCGTAATCC	22	0.0039	0.0039	0.0044	0.0043	0.0040	0.9510	0.7945	1.0163	0.9662	0.9631	1.3566	
ATCACCCCAAACGTCTGTTATCGTAGCCGGTTGTTAGTCT	23	0.0044	0.0044	0.0044	0.0041	0.0038	1.5544	1.3220	1.4430	1.6440	1.5951	1.2240	
CGCTTGATCAGTTCGAATAAATAGCTTGTTGACTGGTGC	24	0.0041	0.0032	0.0045	0.0047	0.0019	1.2901	0.9702	1.3775	1.4470	0.8521	0.7642	
GGGTAGAACAATTCACTTCAGTTTAACTCGAGGTTGTTAG	25	0.0031	0.0027	0.0029	0.0026	0.0025	1.0002	0.9853	1.0542	1.0690	1.1413	1.3065	
CAGGCACAGGTTTGTTCTATTCGTTGACTGACAACGGCA	26	0.0032	0.0035	0.0033	0.0029	0.0023	1.1546	1.2109	1.1147	1.2368	1.2472	1.0306	
CGGCAGGTTAATGGGATAGCAGTCTTGGTATGGTTGTTAG	27	0.0032	0.0033	0.0033	0.0028	0.0033	1.0173	0.7395	1.0216	0.8270	1.0637	1.5280	25
ACACAATGTGCAAGTAACTCGCGAAGTAGAAGGTTGTTAG	28	0.0033	0.0036	0.0034	0.0030	0.0031	1.4062	1.2663	1.3214	1.3078	1.2826	1.2077	29
CACCGTAGTTCAAGCTTGTGCCCGGTAGTTTTGTTGTTAG	29	0.0034	0.0031	0.0032	0.0031	0.0022	1.2203	1.0334	1.3140	1.3988	1.2290	1.2281	
ACCAGGCATGAATAAGTTGGAGATTATGTTTGGTTGTTAG	30	0.0042	0.0045	0.0042	0.0041	0.0033	1.1992	0.9593	0.9814	1.1992	1.1254	1.4056	
GCATGATGTAAGTAACTGGCATTAGGAGTGATGTTGTTAG	31	0.0047	0.0048	0.0047	0.0050	0.0040	1.0891	0.9152	1.0177	1.1970	0.9737	1.1832	
ACAAGGGCATACGTCAGATAGTTTTTATCGTTACTGTCAC	32	0.0047	0.0038	0.0048	0.0052	0.0025	1.1738	0.9408	1.2237	1.3365	1.1448	1.1162	
CATCGGAATCAGTTCTAGCTACGTCGTCATCGTGAAAAGG	33	0.0054	0.0051	0.0056	0.0047	0.0051	1.9018	1.6448	1.7446	1.6493	1.9057	1.4469	17
TAGAGGGTAAGGACGTTAAAACTGTGGGTTGTTAGTCTGG	34	0.0032	0.0029	0.0029	0.0035	0.0022	1.0552	0.8457	0.8081	1.1652	1.0288	1.3692	
ATGCTACACCCGAAATTTGGAGGTTCTTAGTTTCATGCGT	35	0.0024	0.0022	0.0023	0.0022	0.0017	1.0528	0.8443	0.9296	1.0353	0.9966	1.0374	
CCGCTCGCTAACTAAATAGTTCAAAGGCTAGGTCCCCGGA	36	0.0017	0.0014	0.0019	0.0020	0.0013	1.0637	1.0849	1.2386	1.3523	0.9065	1.0532	
ATACCAGATCGTTAGTGTGCAGTCACTTGGAGGTTTTTAG	37	0.0032	0.0030	0.0031	0.0026	0.0025	1.4311	1.5294	1.6862	1.2896	1.6139	1.2754	
CGCGGTCATAGTTCCATTTGAATCGTTGACTGAAGGCTA	38	0.0030	0.0031	0.0032	0.0032	0.0018	1.1678	1.0968	1.2882	1.3737	1.0913	1.2433	
CCATCGTCATCATGAAATAAACGGGTAGGCAACTTCATCC	39	0.0037	0.0040	0.0039	0.0034	0.0030	1.3605	1.2669	1.5108	1.3936	1.5763	1.1260	
CCCACAGCATGTATTATTTAGGAAACCTTTTACTTGGTG	40	0.0019	0.0018	0.0019	0.0018	0.0018	0.7244	0.6418	0.6955	0.6416	0.7001	1.2394	
TCGGCCTACATCTTTGATATGAACGGAGGTTCTTAGTGGT	41	0.0025	0.0024	0.0024	0.0022	0.0020	1.2384	0.9012	0.9926	1.0528	1.5856	1.5776	
ACCAGGTCAGTTTATTCTTCATTTTCGCATTCTGAAGTGG	42	0.0018	0.0018	0.0018	0.0017	0.0016	1.5624	1.3565	1.4093	1.7131	1.6091	1.1853	
CAAGTCACGGAAAGTATCGTTCGTTTGGTCAGGTTGTTAG	43	0.0028	0.0030	0.0028	0.0026	0.0027	0.9397	0.8448	0.8740	1.0189	1.1750	1.2534	30
ACCACGTCAGATATCCATTCGTTCCTCTGAAAGTTCATGG	44	0.0026	0.0027	0.0027	0.0023	0.0021	1.4531	1.3285	1.3039	1.3106	1.3455	1.1281	
CGCCGACATGTAAGTGATTTCAGTAGTCGGTTGTTAGTCA	45	0.0026	0.0026	0.0024	0.0025	0.0016	0.9359	1.0256	0.9720	1.1452	1.0440	0.8952	
CGGCCAGATTTCTTCCTTTTTCGTTAACTGAGGTTGTTAG	46	0.0020	0.0018	0.0018	0.0017	0.0013	1.1170	0.9714	1.0713	1.1995	0.9686	1.1961	
CTGTCCTATCCGGAAATAGTATGTCTTGGGGGGTTGTTAG	47	0.0038	0.0036	0.0034	0.0033	0.0029	1.0238	0.9167	0.9206	1.0042	1.0803	1.3496	
CTCCTCGTCAAGTGAAATTAAAGGACAGTACAGTTTCCGC	48	0.0035	0.0031	0.0035	0.0037	0.0021	1.5131	1.1041	1.2420	1.5512	1.1974	1.3101	
TCACTCCAACTGTAATTTGAACTGTTAGTATGGTTGTTAG	49	0.0036	0.0037	0.0035	0.0032	0.0027	1.6844	1.5653	1.7572	1.7198	1.9290	1.3882	
TCATGGGCAATGTGGGAACCGTAAACTTAGAGGTTGTTAG	50	0.0028	0.0026	0.0028	0.0027	0.0017	1.0944	0.8889	1.0448	1.1982	1.0074	1.1348	
CCACGGTATCCTTCCCTATAGAAATGCTTGTTGATCTGA	51	0.0020	0.0018	0.0019	0.0019	0.0017	1.0951	1.0000	1.0111	1.1781	1.0921	1.2054	
TGTTGACATAGAGAGAACTGTTAAGGGTTGCGGTTGTTAG	52	0.0027	0.0023	0.0023	0.0029	0.0021	1.0996	0.8007	0.8949	1.1082	1.1285	1.3324	
GCGGCTAGTTAGGTGATATTCGTTAATTGGAGGTGCTTAG	53	0.0022	0.0021	0.0022	0.0021	0.0019	1.0387	0.9661	1.1064	1.2052	1.0341	1.1950	
TAGTCGAAGGGAATTACGTCACATGAATATTATGCCATCC	54	0.0060	0.0062	0.0063	0.0057	0.0031	2.9874	2.5890	3.1918	2.7690	2.9506	1.0476	43***
GGCGCTAAGTCTGTTTAGGCATTGAAATTGTTCAGTCTCC	55	0.0015	0.0015	0.0014	0.0013	0.0017	1.9033	1.4327	1.4843	1.4774	1.8105	1.2681	
CAACCCGTCTTGCATAGTCCATAGCTACTCGTCAACCGAA	56	0.0027	0.0024	0.0027	0.0032	0.0016	1.0937	0.9028	1.1367	1.4064	1.1787	1.0581	
CAACAGGACGTTTCGTCTTGGAGGTTTTAGTCTGCCCGA	57	0.0006	0.0007	0.0006	0.0008	0.0006	1.4028	1.3343	1.1599	1.5255	1.4499	0.9713	
ACCACACCGTATCTCTCTATAACTTGGACATACTAGTGG	58	0.0012	0.0012	0.0013	0.0014	0.0012	0.8235	0.7428	0.7879	1.1691	1.0460	1.2823	
ACGCGGACAATCTGTCTTAGTTAGTTCAAGTAGTGCGTGA	59	0.0007	0.0006	0.0007	0.0014	0.0024	0.7670	0.6070	0.7437	1.3284	2.0860	2.9333	37
CCGTCGGCAAGTGAAATAAGTACGTAGTTTATCGTAGACC	60	0.0018	0.0019	0.0017	0.0019	0.0013	1.3612	1.2596	1.2753	1.5685	1.3900	1.1441	
CCGGGCCTATTGACAATTAAACTCTGAGGTTTTTAGTCGG	61	0.0006	0.0006	0.0006	0.0005	0.0006	0.8820	0.8778	0.8891	0.7936	1.1884	1.8394	
CGGCAATATCTCGTTGACTGATTCATTTCTAACGTCCGTA	62	0.0019	0.0013	0.0021	0.0026	0.0015	1.0037	0.7735	0.9414	1.4161	0.9600	1.1761	
ACTCGCTGACTAAAGCTATGTATAAGACGAATCACTCTGG	63	0.0023	0.0021	0.0023	0.0028	0.0017	1.1982	1.1235	1.3755	1.5461	1.5649	1.1838	
CTGCTAATTCGGAACTGACCTGGGGGAGTATGGTTGTTAG	64	0.0020	0.0020	0.0019	0.0018	0.0018	1.0646	1.0631	1.1070	1.4330	1.1591	1.1860	
CCAGCCTTGATAGTCAGATTCTAAGCTTATTGAAATGGGC	65	0.0017	0.0020	0.0018	0.0016	0.0019	2.0322	1.9686	1.8724	1.6754	2.6877	1.5539	
ACCGGCACAAATAGTTATGGAATGCTTGGATGGTTGTTAG	66	0.0013	0.0013	0.0012	0.0011	0.0014	0.9497	1.0452	0.9695	1.2878	1.1834	1.5053	
CCCGGTGAAGATAGGAACATGCAGGTTGTTAGTCATTCGC	67	0.0013	0.0014	0.0015	0.0018	0.0018	0.8389	0.7610	0.7245	1.3309	1.3242	1.7222	

Cluster Parent Sequence	Cluster Rank	Cumulative Cluster Frequency					Mean Cluster Enrichment Ratio						Tested (R14 Rank)
		CRC	Breast	Lung	Prostate	R14	CRC/NH	Breast/NH	Lung/NH	Prostate/NH	R14/NH	R14/Counter	
CAACCACGTAGGGATTGTTAGTTTAAATTTAGCCAAGTGA	68	0.0027	0.0028	0.0030	0.0025	0.0025	1.7732	1.7201	1.6647	1.6500	1.9238	1.5361	32
CACGAACGGAATGCCGGATATTTAGCTGTGTGGTTGTTAG	69	0.0014	0.0017	0.0014	0.0013	0.0018	1.3605	1.0274	1.2294	1.1956	1.5869	1.6223	
GGGGCATGTTCTAAGGTTACCAGTCCATTGAAATACGTCT	70	0.0010	0.0011	0.0011	0.0010	0.0011	1.3461	1.1389	1.0995	1.1644	1.5856	1.4280	
CCTGGGACATGTAATACTTTGGTCCGGTTGTTAGTTTGTT	71	0.0016	0.0016	0.0016	0.0015	0.0015	0.7621	0.7713	0.7222	0.7554	1.0268	1.6489	
ACACAGGTCTGTTCTCTTATTTCGTCATCTGAAGCTACTC	72	0.0010	0.0011	0.0010	0.0008	0.0010	2.1466	1.7032	1.8019	1.5656	1.8973	1.3930	
CATGCGCAGTTCAATAAATCTTGCCTTTGAACTAGTTACC	73	0.0023	0.0024	0.0023	0.0020	0.0023	1.6397	1.4111	1.6238	1.6811	1.9358	1.5015	
CACCTCAGGCACGTTCAATTTGTCATTGAAAATGGTTACC	74	0.0011	0.0013	0.0012	0.0010	0.0013	1.8177	2.2601	1.9907	1.8572	2.4112	1.4418	
CCCTCGTCAACGGAATCATACAAAGTCAGATTAGGACGCA	75	0.0013	0.0010	0.0013	0.0013	0.0008	1.4374	1.1737	1.4462	1.4669	1.3644	0.9538	
ACCCTGGATCACCCTTATCAGTGCAAATTTCGTCAAGTGA	76	0.0016	0.0016	0.0018	0.0014	0.0016	1.6387	1.2789	1.4590	1.3118	1.5656	1.5039	
CCCAGGCAACAGTTCGTCGTTAGCCAATTGAAAATCGGTA	77	0.0008	0.0009	0.0009	0.0007	0.0011	1.4887	1.6851	1.7124	1.4204	1.9088	1.2614	
CTCATGTTCTCCTCAGTAGTATTACTTGGAGGTTCCTTAG	78	0.0012	0.0012	0.0010	0.0011	0.0012	1.3542	1.1115	1.2703	1.3419	1.5062	1.4372	
CAACCAAAACTTGCCATTTGAAATAACTAACCCATCCCGC	79	0.0010	0.0009	0.0010	0.0009	0.0006	2.4346	1.6206	2.0508	1.6702	1.5121	1.1747	
CGATCCGGAACACTAGGAATTTGGAGGTTGATAGTCTGTT	80	0.0004	0.0003	0.0004	0.0004	0.0005	1.1024	0.8117	1.0360	1.3329	1.2038	2.1361	
CAGCCAGGTGAAAGTTAAAGTCTGTTGATTGTTACCGACC	81	0.0012	0.0013	0.0012	0.0011	0.0013	1.5332	1.3225	1.0594	1.3726	1.4743	1.2387	
CGCCAAGTGAAATTGAAAGAGATTAAGCCTCAGTTTCGCC	82	0.0014	0.0014	0.0015	0.0014	0.0011	1.2003	1.1499	1.3428	1.3403	1.1688	0.9879	
CGGCATCATGAAACTGAAGTGCTTTACCCGTTCGTCATCC	83	0.0015	0.0016	0.0015	0.0014	0.0013	1.6012	1.6196	1.4514	1.7342	1.9945	1.6547	
CGTGGAGGCAGAAGATTATACCTCGTAAACTAGTGATTCGG	84	0.0005	0.0006	0.0004	0.0004	0.0007	1.0880	1.1909	1.0146	1.2112	1.7162	1.5081	
CAGCGCTAACTGAAAGCCTTTATTGAGTCATATCCAGGCA	85	0.0010	0.0010	0.0010	0.0012	0.0012	1.1803	1.1417	1.0490	1.1791	1.6376	1.4420	
TGCCTGTTCATGTATACTTTTCAGGTCGTTGGTTTTTAG	86	0.0009	0.0009	0.0009	0.0007	0.0008	1.2923	1.2834	1.3645	1.2929	1.2790	1.1494	
CCGAGAGATCATGTTCCATCTATATTTGTCAACTGGCAA	87	0.0014	0.0012	0.0014	0.0013	0.0013	1.1993	1.0174	1.2229	1.1980	1.4406	1.3617	
ACCACCAAGTCAGGTTTTGATCCGTTTGTCGACTGAAGCC	88	0.0009	0.0010	0.0010	0.0009	0.0006	2.4767	2.3662	2.6167	2.2265	1.5511	1.0477	
TGCCATCCATCAGGCTCTTGCTCAAGTAGAGGGTTGTTAG	89	0.0010	0.0011	0.0010	0.0010	0.0010	1.1640	1.2370	0.9956	1.4028	1.1053	1.2049	
CCGGCAAGTGAAATGCTGTGGAAATCTGTTTAGGCTATCA	90	0.0010	0.0011	0.0010	0.0011	0.0010	0.9136	1.0207	1.1418	1.2985	1.4368	0.9643	
CACCCGGTCATAGGTATTTACATCGTCGACTGACATAGGT	91	0.0010	0.0008	0.0009	0.0011	0.0007	1.6648	1.1806	1.1492	1.4539	1.5696	1.7072	
CATCGCTAACTGAGTAAATACAAGGTCAATCAACCTGCAG	92	0.0017	0.0014	0.0016	0.0020	0.0008	1.5483	1.1404	1.3433	1.8589	1.0899	0.8574	
TCGCCAAACACCGTGCTACATTGGTTTGGTTCTTAGTGGA	93	0.0012	0.0012	0.0012	0.0012	0.0008	1.3830	0.9970	1.1349	1.2420	1.1207	0.9334	
CCCACATAAAGTCGAAGTTCATAATCCCTCGTTTTGTGAA	94	0.0012	0.0010	0.0012	0.0009	0.0011	1.6175	1.2622	1.5000	1.1751	1.8593	1.7191	
CTTCATGAGCGGTAGTAGCTTGGAGGTTTTTAGTTGGTGA	95	0.0007	0.0007	0.0007	0.0006	0.0008	1.0551	0.9929	1.0752	1.1669	1.2843	1.8299	
ACCCACAGTGAAGTAATATCGCTAACTGATAAGACTTTGT	96	0.0017	0.0015	0.0019	0.0018	0.0014	0.9350	0.9460	1.1539	1.0933	1.0470	1.2309	
CCCCGAGTCGAAAAGGTAGTCAATGTTCTTTGCCAAGTGA	97	0.0008	0.0007	0.0008	0.0006	0.0008	1.2494	1.0804	1.1503	1.2167	1.6377	1.1524	
CGATCACGTGTTCATAACAAGACTTCTTGCTGACTGATGT	98	0.0013	0.0011	0.0012	0.0014	0.0008	1.5792	1.2127	1.3819	1.7392	1.2927	1.0777	
CGATCGGTAACTGAAAAGAGATGTTTGTCGTTAGGTTCC	99	0.0015	0.0014	0.0015	0.0017	0.0013	1.2525	1.1457	1.3979	1.4104	0.9946	1.0939	
GCAGCCCTATTTCACAAGACTTGGAGGTTGTTAGTTAGAT	100	0.0004	0.0004	0.0004	0.0004	0.0004	0.9168	1.0104	0.8888	1.0036	1.1470	1.3504	

Figure 7. PanC2 Selection Sequencing Data. Target Selectivity by Cluster for Top 100 clusters. Cumulative Cluster Frequency = Mean Cluster Frequency, by cluster. Target enrichment ratio, by cluster.

Beginning with cluster 1, we note the mean cluster enrichment ratios of the counter selection cancer types relative to normal human serum hover close to a value of 1, indicating that cluster 1 sequences are not particularly selective for any of the counter selection cancers compared to normal serum. However, when comparing the ratios of pancreatic cancer to normal human serum and to the counter selection cancer pool, the ratios are 1.29 and 1.22, respectively. These values indicate preferential selectivity for pancreatic cancer over the other targets compared, translating to cluster 1 exhibiting pancreatic cancer specificity.

Other notable clusters that warrant highlighting are Clusters 54 and 59. Cluster 54 demonstrates significant selectivity for all cancer types against normal human serum with high magnitude of comparison (ratio range 2.5 – 3.2), indicating significant potential for inclusion into a diagnostic assay for general cancer detection. Unfortunately, Cluster 54 appears to lack the specificity

required to distinguish pancreatic cancer from other cancer types (ratio 1.05). Consequently, its use would be limited to a preliminary cancer screening test. Conversely, Cluster 59 shows minimal selectivity towards the counter selection cancer types compared to normal human serum, but demonstrates significant selectivity/specificity for pancreatic cancer against both normal human serum (ratio 2.1) and other cancer types (ratio 2.9).

Based on the data displayed in Figure 7, four contending clusters (54, 59, 61, and 80) with notable specificity for pancreatic cancer have been selected for further exploration of their individual sequences using Log-Log Sequence Frequency Plots, as displayed in Figure 8.

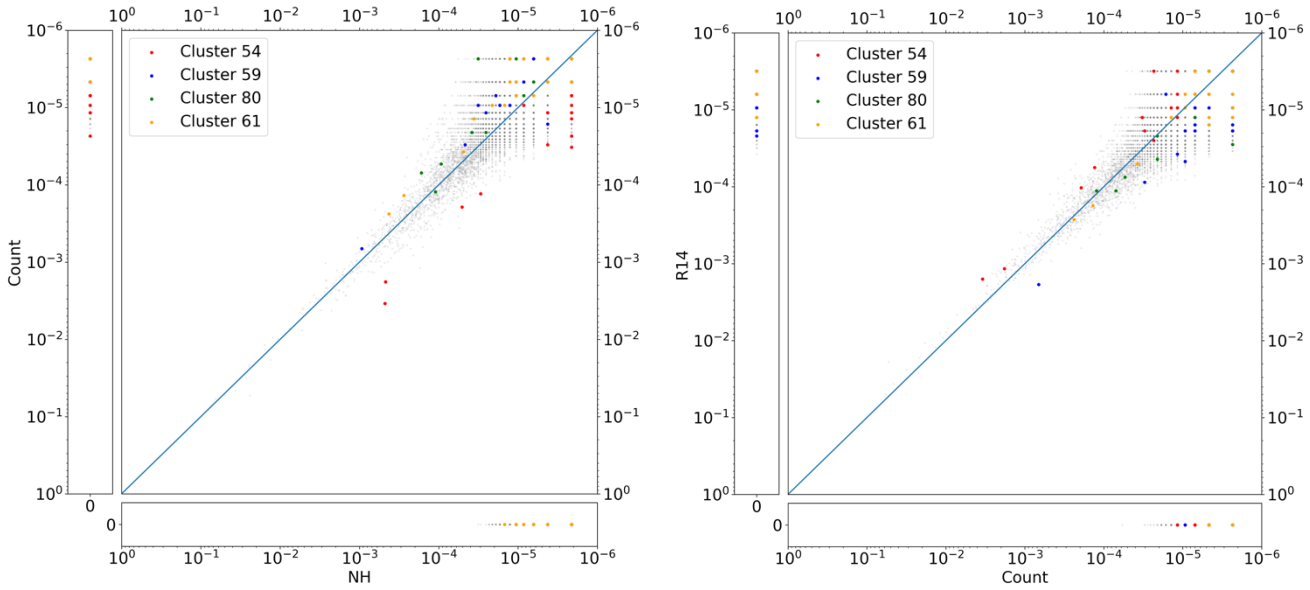


Figure 8. Log-Log Sequence frequency plots, by target cluster. Left: Compare Counter Selection Target Cancers vs. Normal Human Serum. Right: Compare Pancreatic Cancer Serum (R14) vs. Counter Selection Target Cancers.

Figure 8 reinforces the conclusions from Figure 7, particularly reinforcing Cluster 54 as the most promising differentiating cluster for both pancreatic cancer and other cancer types against normal serum. Additionally, Cluster 59 remains most capable of distinguishing pancreatic cancer from other cancer types. Based on these projections, the cluster representative of Cluster 54 was later added to the sensitivity and specificity analysis of the Pancreatic Cancer Selection section (Chapter 2). The challenge in adequately evaluating the cluster representative (Rank 43 in the top unique sequences by frequency at Round 14), is the lack of access to sufficient patient samples to perform a high-powered sensitivity and specificity analysis. Nevertheless, the results of the sensitivity and specificity analysis of 14-43 are depicted in Figure 9 along with a discussion of the correlation of the predicted performance to the actual cleavage patterns of 14-43 in real patient samples.

The results displayed in Figure 9 suggest that Cluster 54 (14-43) may indeed show preferential cleavage towards cancer serum samples, as evidenced by the Healthy Participant Serum yielding the lowest cleavage values with the exception of the Fresh Collected Pancreatic Cancer serum samples. The caveat of the fresh collected samples is the extreme degree of non-specific nuclease degradation of the DNAzyme probes casting doubts on the accuracy of the calculated cleavage percentage. Nonetheless, all OTB cancer samples and the fresh collected pancreatic cancer sample similarly seroconverted by defibrinating plasma show preferential cleavage of 14-43. These results indicate that 14-43 is a real contender as a general cancer detection probe pending additional testing and optimization.

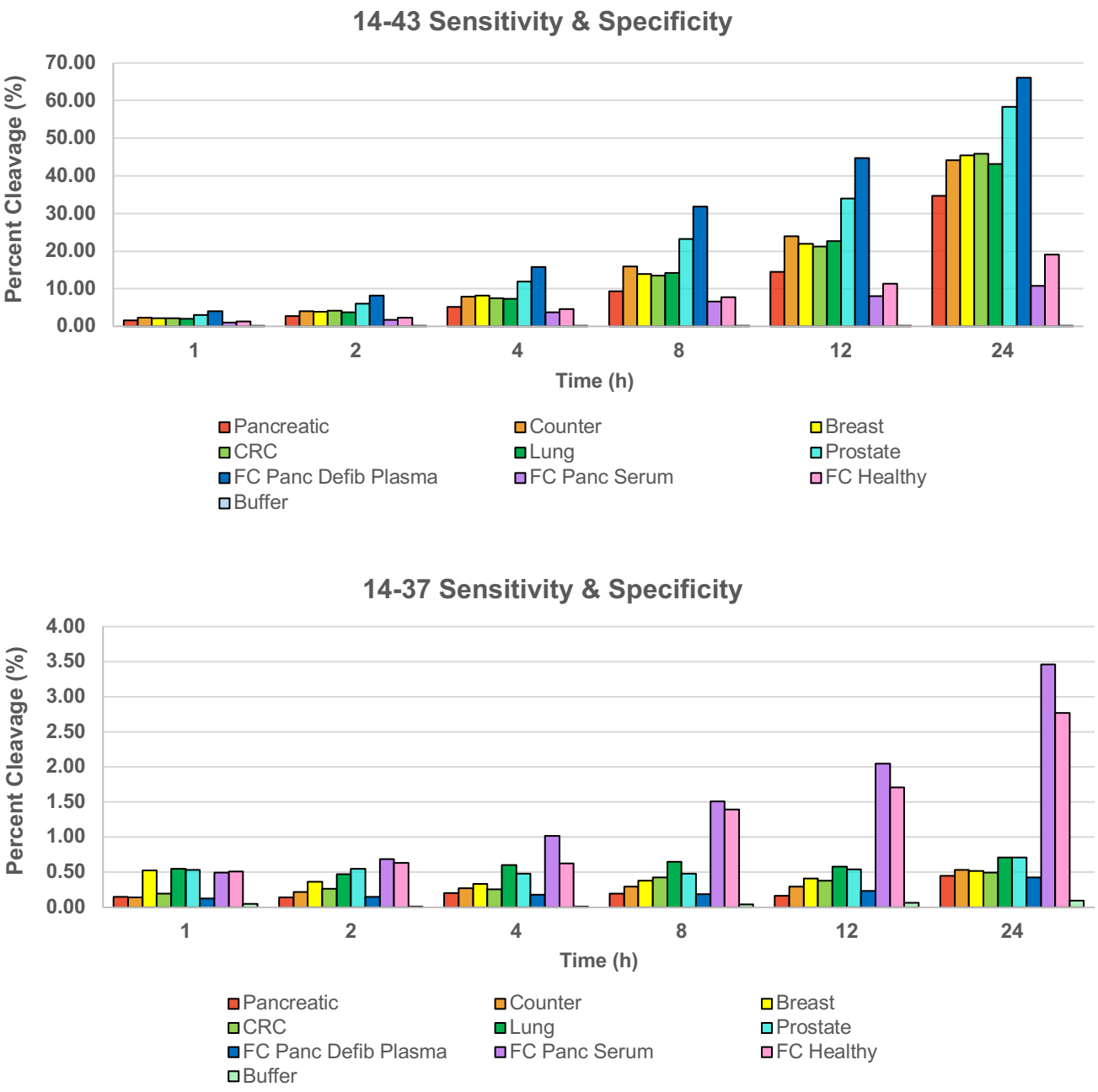


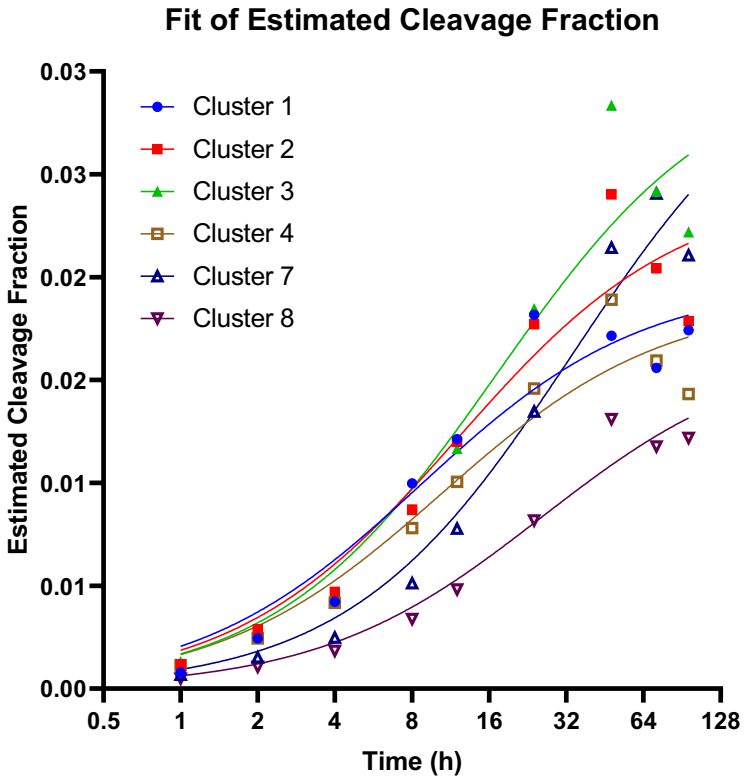
Figure 9. Percent cleavage values of time course reactions with the cluster representative sequences of Cluster 54 (14-43) and 59 (14-37). Counter = Counter Selection Cancer Pool. CRC = Colorectal Cancer. FC Panc Defib Plasma = Pancreatic Cancer Serum obtained by Defibrinating Fresh Collected Plasma. FC Panc Serum = Fresh Collected Pool of Pancreatic Cancer Serum. FC Healthy = Fresh Collected Healthy Participant Serum.

Conversely, Figure 9 shows that 14-37, the cluster representative of Cluster 59 does not appear to show preferential cleavage of pancreatic cancer samples, with the exception of Fresh Collected Pancreatic Cancer serum samples. However, increased cleavage is also noted in the Fresh Collected Healthy Participant serum, suggesting the cleavage may be associated with the method of serum conversion or the recency of sample collection and possibly associated higher nuclease activity. Nonetheless, through all six time points, the rate of cleavage of 14-37 in Fresh Collected Pancreatic Cancer serum exceeds that of Fresh Collected Healthy Participant Serum, suggesting 14-37 may warrant additional testing and optimization.

Having concluded an in-depth extraction of target-oriented data for identification of sequences displaying maximum sensitivity and specificity for pancreatic cancer and other cancer types, we subsequently turned our attention to the identification of fast-acting clusters based on kinetic prediction analysis from the cleavage time course sequencing data. This analysis is presented as a proof of concept for the degree of data extraction that can be accomplished from cleavage time course sequencing data. However, when applied to serum in the absence of a method for controlling nuclease activity, the data will suffer from increasing errors over time due to the introduction of multiple competing cleavage rates.

Figure 10 depicts the results of the predicted kinetic properties of the six fastest-acting clusters. The DNAzyme reaction rate modelled uses a first order reaction approximation – specifically a pseudo-first order model to predict expected cleavage fractions for each of the six clusters over time. A pseudo-first order model is used given the assumption that the cleavage rate depends on the concentration of DNAzyme in the reaction. To fit the model, we plotted the mean frequency

of each cluster across the nine time points submitted for sequencing to fit the pseudo-first order model. Subsequently, the plateau (Bmax) of each predicted model reflects the frequency of each cluster in the total Round 14 population. The relevance of this caveat is to differentiate the plateau values in this predicted model from the values we expect to see in a pure experiment with individual sequences. In the latter, the individual sequence being tested represents 100% of the DNA in the reaction and consequently the plateau will be higher. Based on the modelled data, rate constant K_d values are provided.



	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 7	Cluster 8
Best-fit values						
Bmax	0.01978	0.02437	0.03056	0.01895	0.03246	0.01667
Kd	8.619	12.08	17.10	10.42	33.74	25.66
95% CI (profile likelihood)						
Bmax	0.01646 to 0.02385	0.01960 to 0.03077	0.02420 to 0.04004	0.01555 to 0.02328	0.02594 to 0.04361	0.01358 to 0.02146
Kd	4.481 to 16.33	5.820 to 24.97	8.308 to 36.28	5.253 to 20.39	19.03 to 64.72	14.55 to 47.71

Figure 10. Fit of estimated cleavage fraction for 6 clusters based on sequencing the cleavage time course of R14 population vs. Pancreatic Cancer Serum.

As can be seen in the raw predicted kinetic data, the confidence intervals of the rate constant k_d are quite wide, indicating the model is prone to errors, most likely due to the introduction of additional background rates over time caused by non-specific nuclease degradation. This effect is particularly evident in the latter time points through the signal decay observed in the plateau of the modelled cleavage reactions. Therefore, while we cannot certify the values of the predicted rate constants, they can still be used to offer a relative ranking of clusters. For instance, based on the values of the rate constants, the clusters can be ordered in descending order as Cluster 7 > 8 > 3 > 2 > 4 > 1, indicating that Cluster 7 would likely hold faster-cleaving sequences in pancreatic cancer than the other clusters, while Cluster 1 would likely hold the slowest-cleaving sequences in the six clusters compared.

The predicted kinetic data generated was then cross-referenced with the analysis performed on 14-2 (Cluster 4) in Chapter 2. Figure 11 is a composite of the kinetic analysis performed on a pure experiment with 14-2, using the results obtained from multiple cleavage time course reactions with every target available. The plateau and rate constant values are provided for the same pancreatic cancer pool employed to generate the predicted data above. While the rate constant data and plateau from the pure experimental data differ from the predicted models, it is interesting to note a similar morphology of the pancreatic cancer pool reaction models over an identical time period, suggesting a sound relative approximation of the real data by the predicted models, particularly as it reflects the introduction of background non-specific degradation rates.

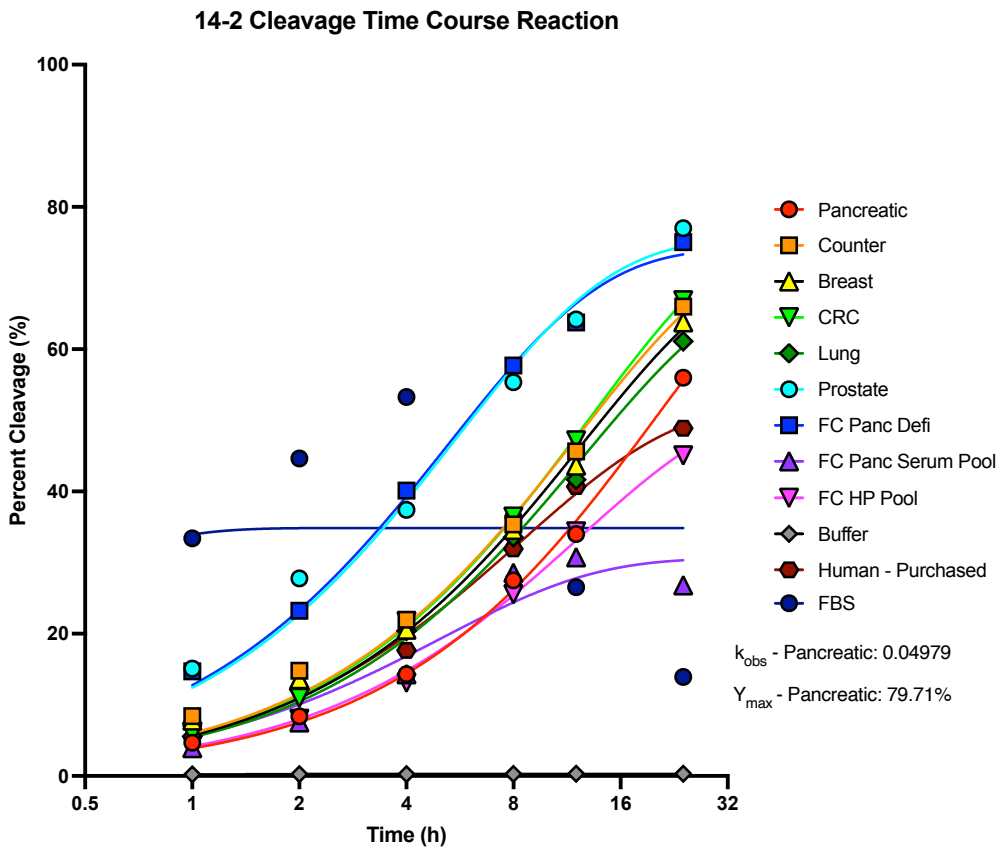


Figure 11. Kinetic analysis of 14-2 cleavage time course reactions in multiple serum categories. Pure experimental data.

In conclusion, the advanced bioinformatic analysis completed in this section has proven effective in the diagnostication and prognostication of the selection experiment conducted in Chapter 2. Much like a radiological imaging scan, we have demonstrated that deep sequencing analysis can offer a cross-sectional look at the progress of a selection experiment, and predict not only the success of the selection but also offer viable candidate sequences with quantifiable selectivity metrics, that would have otherwise been missed using conventional candidate DNAzyme selection approaches. We will attempt to replicate this positive effect by applying the same advanced bioinformatic analysis method to the malignant hyperthermia selection.

4.2.2. Malignant Hyperthermia

4.2.2.1 Analysis design/plan

The malignant hyperthermia bioinformatic analysis follows a similar approach to the pancreatic cancer analysis, with minor changes reflective of the *in vitro* selection protocol. For instance, given the degree of enrichment demonstrated on gel analysis from round to round, we did not feel the need to perform multiple sequencing runs during the *in vitro* selection protocol, instead waiting until a tapering was observed by Round 12 to sequence the enriched libraries from Rounds 6-12. Another notable difference is inherent to the targets being pursued in the malignant hyperthermia selection. Unlike the pancreatic cancer selection which aims to differentiate pancreatic cancer from multiple other cancer types, the malignant hyperthermia selection aims to differentiate MH susceptibility from MH negative samples. Consequently, in the absence of numerous counter selection targets, we were able to prepare a more thorough sequencing run, by performing two parallel cleavage time course reactions with the positive and negative samples, and submit each recovered DNAzyme fraction for sequencing. The effect of this endeavour is the ability to compare MHS vs. MHN sequencing results at each time point, and ideally identify an optimal time point of comparison between the samples, in addition to identifying a promising DNAzyme candidate capable of detecting MH susceptibility. As such, the MH bioinformatic analysis generated is the result of a single comprehensive sequencing run. Yet, this single run offers the same degree of comprehensiveness in the evaluation of the success of the MH selection through an account of the DNAzyme library's enrichment and selectivity over the rounds of selection, as well as a thorough evaluation of sequence clusters' selectivity and specificity, towards identification of candidate DNAzymes capable of MHS detection.

4.2.2.2 Evaluation of DNAzyme library enrichment and selectivity

We begin once again with an evaluation of the *in vitro* selection protocol efficiency by offering an overview of the evolution of the DNAzyme library through the rounds of selection sequenced, specifically Rounds 6 through 12. By examining enrichment patterns across the latter seven rounds of selection, we can infer whether the selection is responding to the selection pressures applied and effectively advancing towards the identification of prominent sequences sensitive and specific for MH susceptibility. Following the same equation for calculating the Enrichment Ratio for each round of selection, we generated a graph tracking the MH1 DNAzyme library's enrichment, illustrated in Figure 12. These results demonstrate steady enrichment of the MH1 DNAzyme library over the course of the selection protocol, matching expected values based on the gel analysis. The enrichment ratio increases over the latter seven rounds of selection from 1.2 in Round 6 to 2.6 in Round 12.

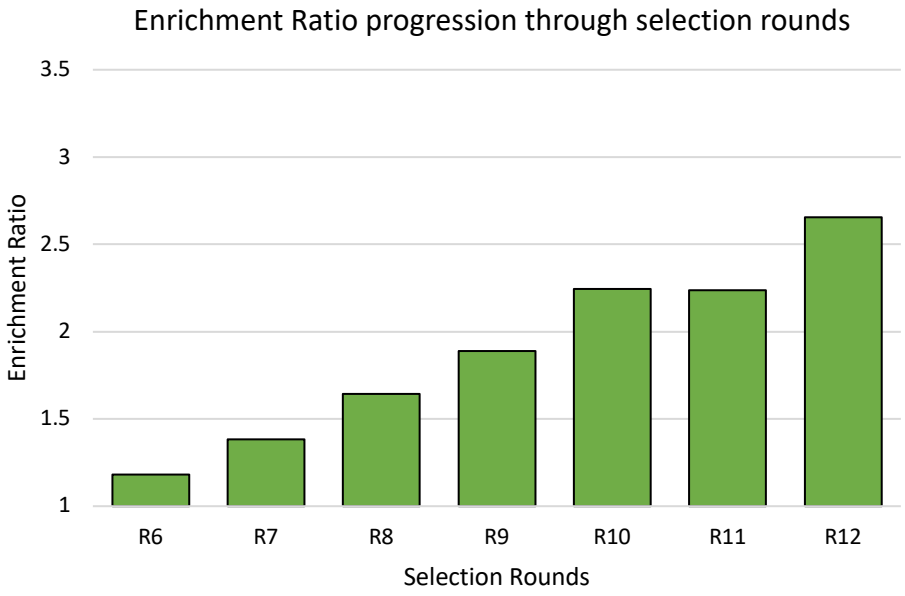


Figure 12. Depiction of round-to-round enrichment over the course of the *in vitro* selection.

We subsequently opted to take a closer look at the round-to-round enrichment through unique sequence frequency measurements in order to better track how to the DNAzyme library is responding to the selection pressures applied, and concurrently identify any high-frequency sequences showing high selectivity for MH susceptibility. Log-Log sequence frequency plots were once again utilized to visualize the data generated, and are displayed in Figure 13.

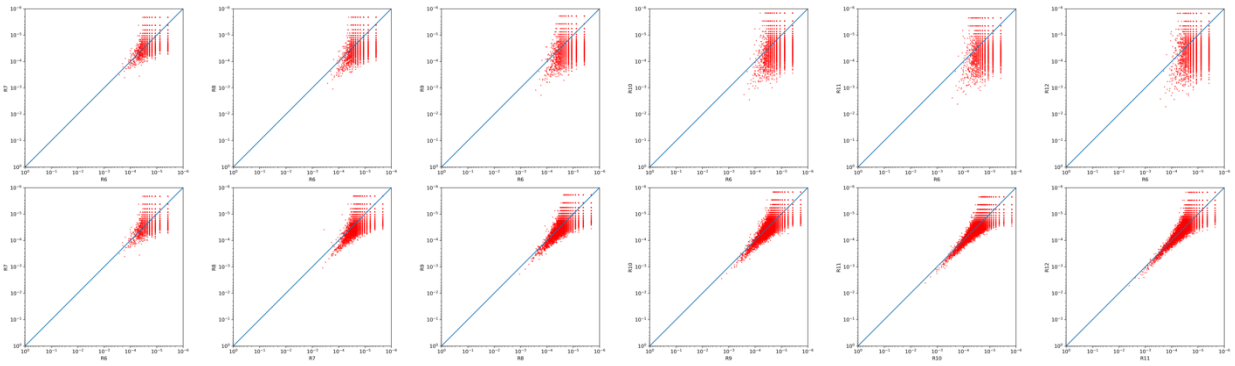


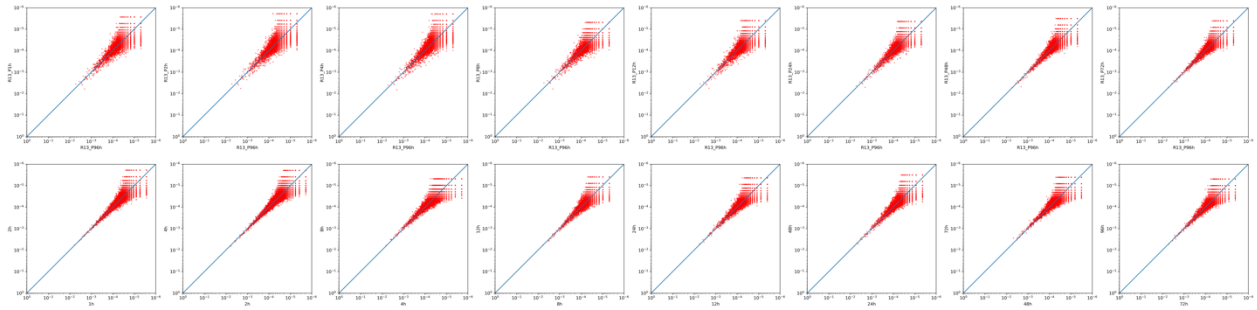
Figure 13. MH1 Selection – Sequencing Data – Population Progression Log-Log Sequence Frequency Plots. **A.** Top Row - Comparison of Rounds 7-12 populations to Round 6. Populations are displaced below the diagonal in each comparison indicating enrichment of sequences. **B.** Bottom Row – Comparison of sequential rounds R7 v. R6 to R12 v. R11. Populations are displaced below the diagonal on all comparisons, except

The top row of Figure 13 illustrates round-to-round unique sequence frequencies for each enriched population compared against a common baseline – the Round 6 enriched library. A gradual downward shift of the sequences is noted over the course of the selection, through Round 12. This trend is observed in high-frequency sequences as well, and indicates that the library is successfully enriching towards a target of interest. The bottom row of Figure 13 illustrates round-by-round unique sequence frequency comparisons, which substantiate the data of the top row, concluding that steady library enrichment is occurring. Once again, the high-frequency sequences and the bulk

of the library continue to fall below the diagonal through Round 12, indicating enrichment is still occurring in the terminal selection round.

Following analysis of enrichment across the latter seven rounds of selection, we turned our attention to the enrichment of the Round 13 library across multiple time points, in an effort to identify any fast-acting sequences. This data was acquired by ligating the enriched library obtained following the twelfth round of selection with the FQ30 signal generating segment and performing a cleavage time course reaction in MHS serum and MHN serum. Figure 14 depicts the results of the MHS analysis. Once again, individual dereplicated sequences have been assigned corresponding frequencies based on their number of detections in each time point, and Log-Log sequence frequency plots have been employed to display the data generated. Similar to Figure 13, the top row represents multiple time point distributions across a common baseline (96h incubation time point), while the bottom row depicts timepoint-by-timepoint comparisons. Additionally, we opted to include a 1h vs. 96h time point comparison of the top 5 clusters in an early effort to identify any fast-acting sequences showing preferential selectivity for MH susceptibility.

A



B

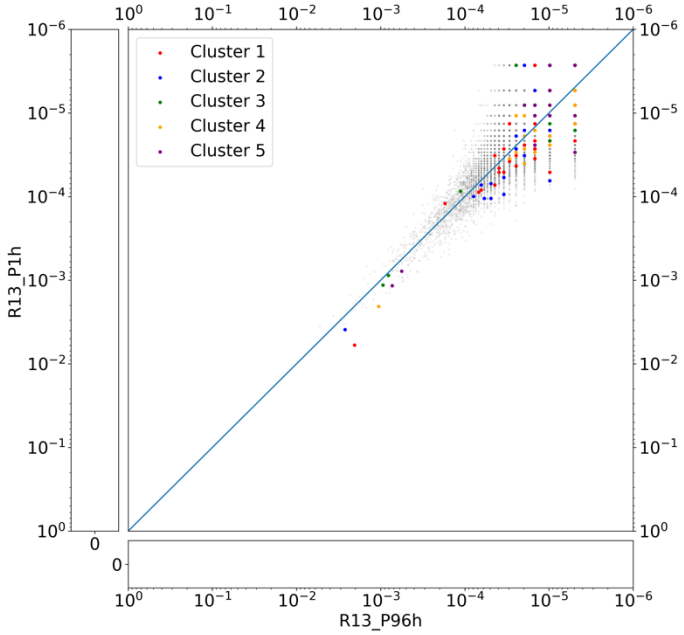


Figure 14. A. MH1 Selection – Sequencing Data –Timepoint Comparison Log-Log Sequence Frequency Plots. Top Row - Comparison of timepoints 1-72h vs. 96h. Bottom Row – Comparison of sequential timepoints 1h v. 2h to 72h v. 96h. **B.** 1h vs 96h time point comparisons of most abundant clusters.

The results of the time course sequencing analysis are similar to those of the pancreatic cancer analysis. Namely, a large data spread can be noted when comparing early time points with the 96h time point. This observation is expected and indicates wide variation of sequence frequencies in the populations recovered from early time points compared to the 96h time point. However, unlike

the pancreatic cancer analysis where more data spread was noted in the 1h vs 2h comparison, the MH analysis shows a compacted data spread for the first two sequential time point comparisons, with a gradual widening of the data spread in the 4h vs. 8h, 8h vs 12h, and 12h vs 24h plots. The accumulated rate differences in the mid-time course time points suggest the selection pressures applied to the MH1 library did not favour the enrichment of fast-acting sequences, and are likely reflective of the incubation times chosen for positive (4h) and counter selection (24h). This finding is consistent with the sensitivity and specificity analysis performed in Chapter 3, in which two distinct sequences showed favourable sensitivity and specificity values at the 24-hour incubation time point.

Figure 14B additionally confirms that the five most abundant clusters from Round 12 hold the fastest-acting sequences in the library, since the majority of sequences within all 5 clusters, and particularly their highest-frequency sequences all fall below the diagonal line. This observation suggests that Clusters 1-5 may hold preferential selectivity towards MH susceptibility.

We subsequently proceeded to compare the counter selection samples' time course results, and once again plotted dereplicated sequences' frequencies on Log-Log plots, identical to the positive selection samples/ time course results. The results are illustrated in Figure 15. We opted to once again explore the performance of the 5 most abundant clusters from Round 12, and compare them to their performance in the positive selection data.

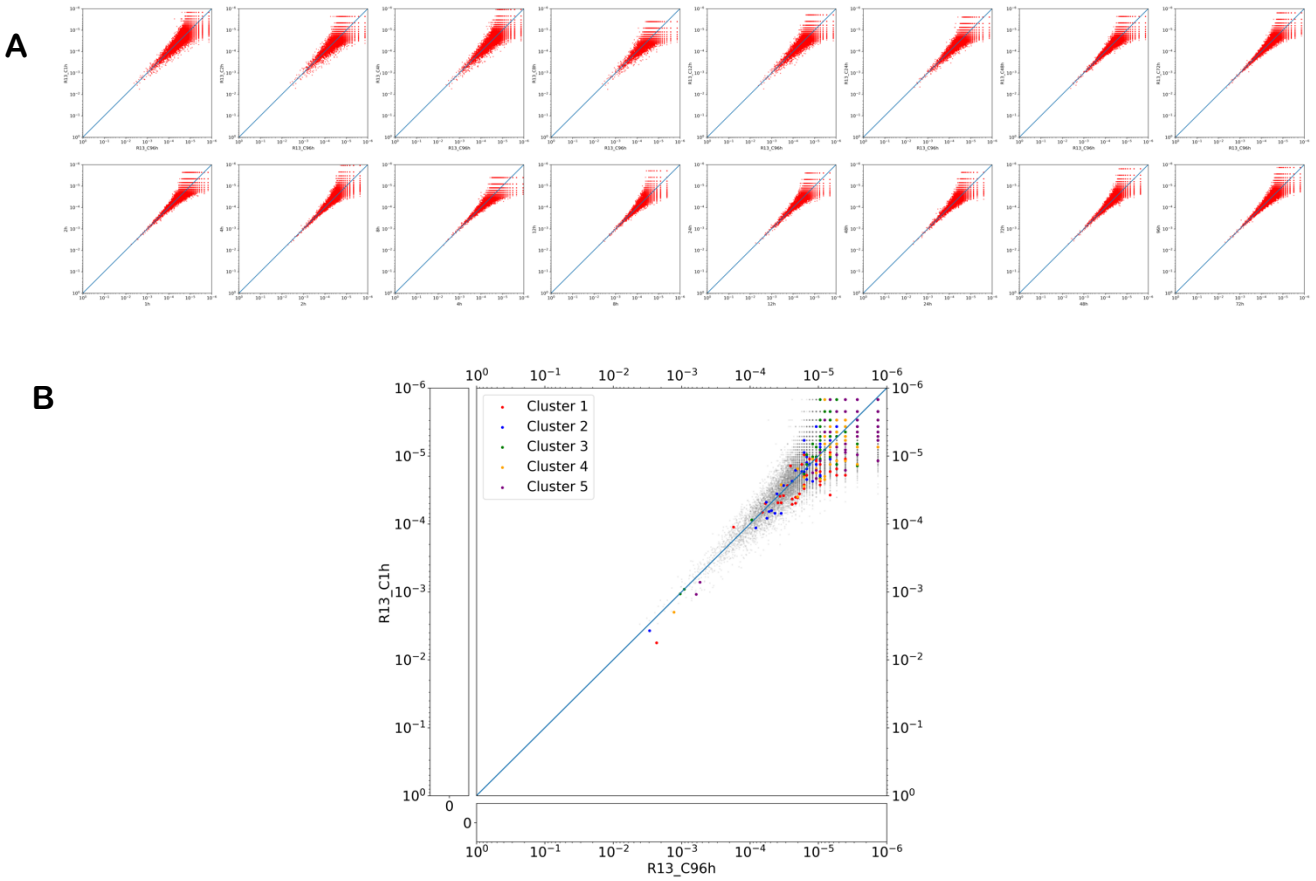


Figure 15. A. MH1 Selection – Timepoint Comparison – Counter Selection Log-Log Sequence Frequency Plots. Top Row - Comparison of timepoints 1-72h vs. 96h. Bottom Row – Comparison of sequential timepoints 1h v. 2h to 72h v. 96h. **B.** Log-Log Sequence Frequency Plots depicting 1h vs 96h time point comparisons of most abundant clusters.

The counter selection time course analysis essentially mirrors the plotted patterns of the positive time course analysis, but with seemingly higher density of sequence reads. This suggests a possibly higher read depth in the counter selection samples compared to positive selection samples, and will be explored later in this chapter. Unfortunately, the log-log sequence frequency plots demonstrating the same trends in counter and positive selection time courses indicate a lack of target specificity. This effect is further observed in the top five clusters from Round 12. Ideal circumstance would show the majority of the sequences in the clusters along or above the diagonal in counter selection, indicating preferential selectivity in positive selection. However, it appears

the cluster sequences have identical positions in both positive and counter selection for the same time point comparison.

Having explored the time dimension of the time course sequencing data, we next turned our attention to the target dimension. By comparing the sequences obtained at identical time points between the counter and positive selection samples, we gain access to a deeper look at potential sequences that show differentiating capabilities and simultaneously identify the optimal time point for differentiation. The results of this analysis are illustrated in Figure 16. Once again, the data was plotted on Log-Log Frequency Plots, followed by a comparison of the top five clusters in MHS vs MHN serum at two different time points, namely 1h and 96h.

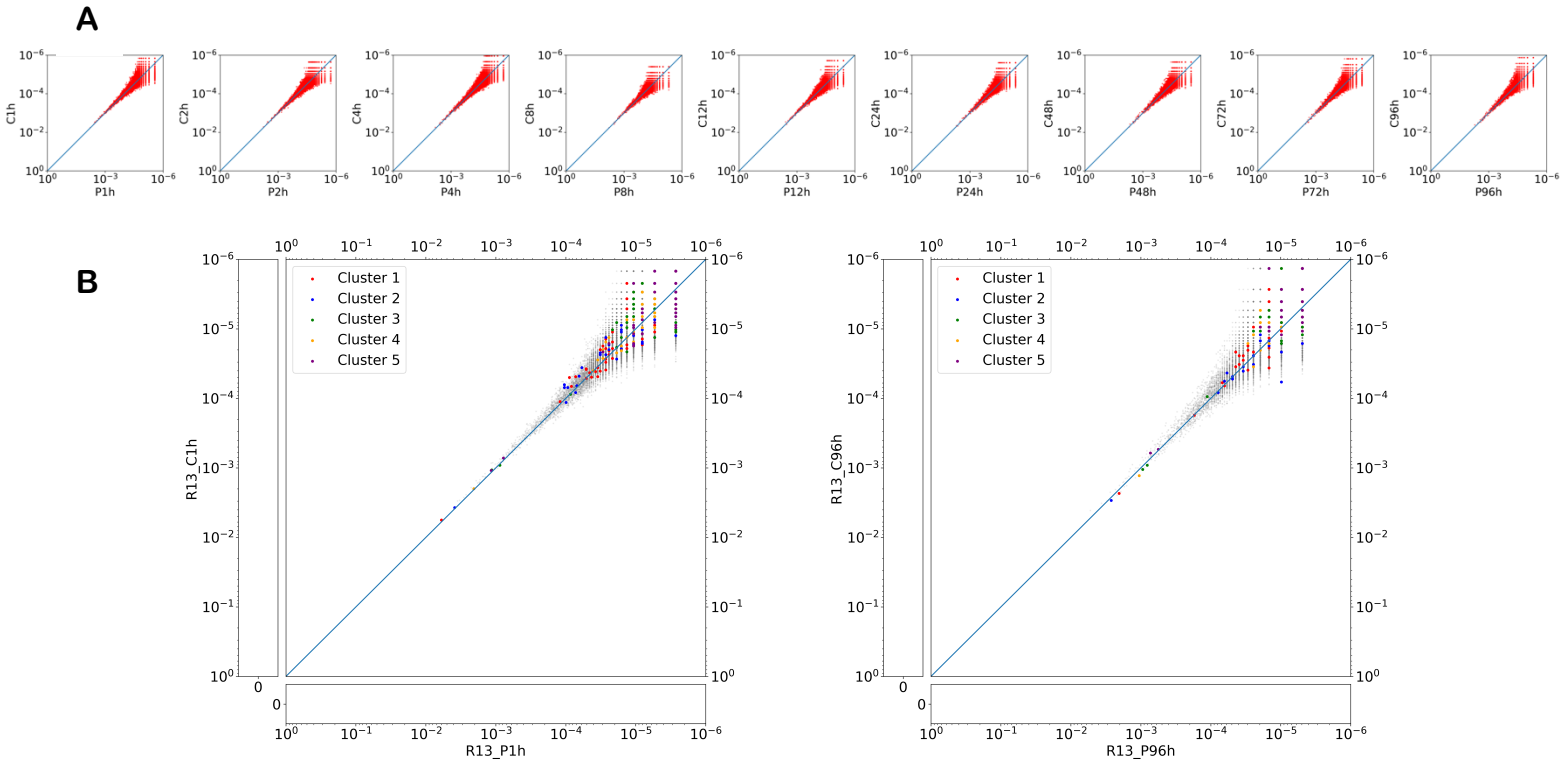


Figure 16. MH1 Selection – Sequencing Data – Target Comparison. **A.** Log-Log sequence Frequency Plots showing comparison between Positive and Counter selection samples at matched time points. **B.** Log-Log Sequence Frequency Plots comparing MHS vs. MHN performance of most abundant clusters at time point extremes.

Figure 16 is a depiction the selectivity of the MH1 library recovered from Round 12. By plotting the positive selection sequencing results along the x-axis, we are searching for sequences that deviate upwards from the diagonal, suggesting preferential selectivity in positive selection samples. However, Figure 16A shows a lack of deviation from the diagonal across all time points. This observation is echoed in Figure 16B, whereby even the most abundant clusters' sequences fall right on the diagonal line at both time point extremes.

Overall, our evaluation of the enrichment sustained by the MH1 library suggests that while enrichment is indeed occurring towards a target, this target is likely not specific to the MH sensitive samples, leading to a lack of selectivity displayed by the MH1 library towards the positive selection samples. Consequently, in the absence of any stand-out sequences emerging from the library enrichment analysis, we turned our attention to cluster analysis in the hopes that a systematic approach to cluster-by-cluster comparison will yield promising DNAzyme candidates capable of detecting MH susceptibility.

4.2.2.3 Evaluation of sequence clusters for identification and projection of promising candidate DNAzyme sequences

We began our cluster analysis by once again employing a sequence aligning algorithm to cluster all recovered sequences from the Round 12 library based on 90% identity, permitting a maximum of four nucleotide mutations in the random domain. Following the formation and identification of the clusters, we proceeded to calculate each cluster's mean cluster frequency and mean cluster enrichment ratio of positive to counter selection samples. Based on the results from the enrichment analysis above, we expect the majority of the mean cluster enrichment ratios to hover around 1,

given minimal displacement of sequences above or below the diagonal line in the Log-Log sequence frequency plots. Figure 17 depicts the mean cluster enrichment ratios of the top 1000 clusters of sequences for each time point collected in the parallel positive and selection time courses submitted for sequencing.

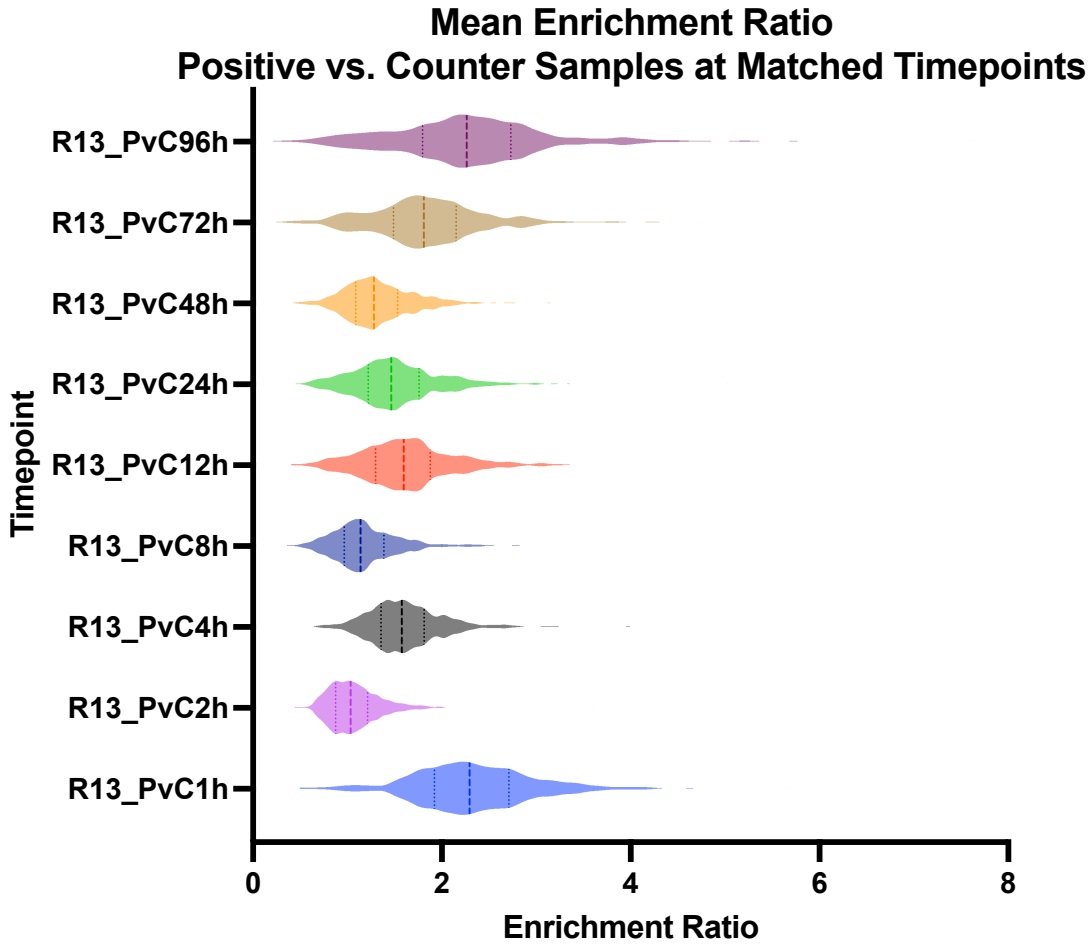


Figure 17. Mean enrichment ratios of positive to counter selection samples for each time point, across top 1000 clusters. The mean ratio value for each time point is denoted by the dashed line.

Figure 17 curiously shows significant deviations of the mean cluster enrichment ratios from the value of 1 across multiple time points, contradicting the minimal changes previously observed in the populations. Given the seemingly contrasting data from the enrichment analysis and this cluster

analysis, we opted to explore an observation made when plotting the counter selection log-log sequence frequency plots – namely the observation of seemingly higher sequencing depth observed in counter selection time point data compared to positive selection data. Consequently, upon comparing the total read counts of the positive and counter selection samples for each time point, we confirmed our hypothesis and established that the disparate sequencing depth differences have a confounding effect on the mean cluster enrichment ratios. Figure 18 illustrates the sequencing depth comparison, illustrating a higher read count at nearly every time point in the counter selection target populations. By virtue of the counter target data having higher read depths, more low frequency sequences are introduced in the population, overall decreasing mean cluster frequencies in the counter target time point data, and ultimately skewing the mean cluster enrichment ratios when dividing the positive target mean cluster frequencies by the counter target mean cluster frequencies.

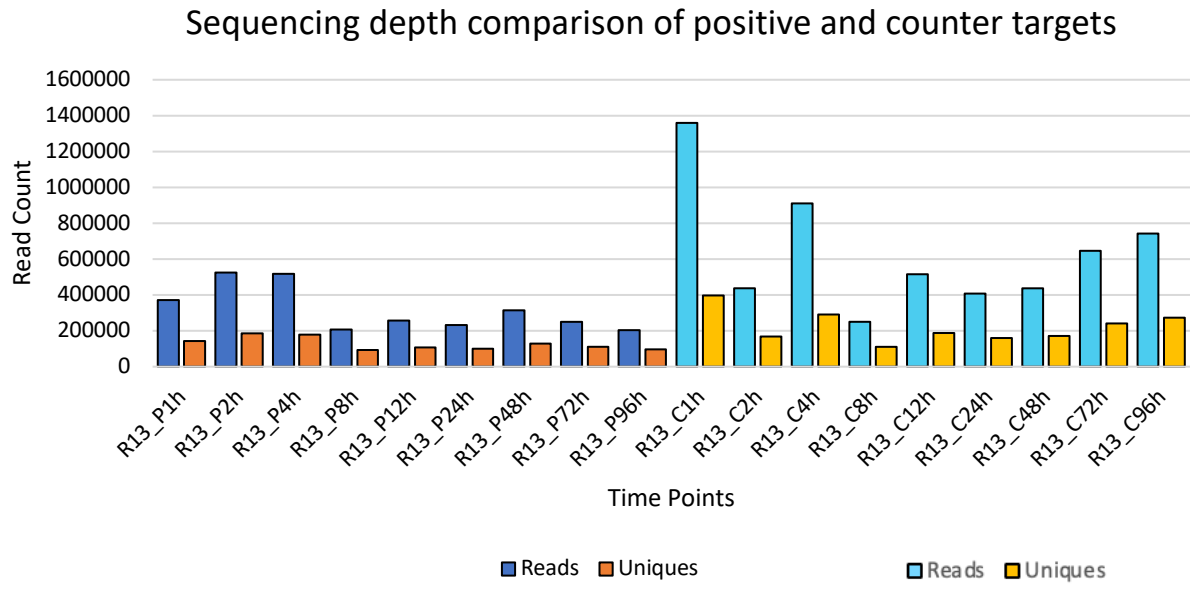


Figure 18. Comparison of total and unique read counts for positive and counter targets across all time points sequenced. P = positive selection target. C = counter selection target.

Consequently, in order to accurately compare the positive and counter target populations at each time point, we proceeded to standardize each cluster enrichment ratio against the population mean enrichment ratio (dashed line value in Figure 17) for each time point population. Specifically, we proceeded to calculate the Z-score of each cluster's mean enrichment ratio by subtracting it from the mean population enrichment ratio. In doing so, we obtained the degree of deviation of each cluster from the mean enrichment ratio of the population (for each time point), allowing for normalization of time point data. Once all the data was normalized against a mean value of 0, we ordered the clusters based on the highest mean cluster enrichment ratio Z-score at the 1h time point, with the top 20 clusters illustrated in Figure 19. The 1h time point was selected given the assumption that the DNAzyme cleavage occurring at earlier time points is more likely to depict intentional target-mediated cleavage as opposed to non-specific degradation of the DNAzymes.

Parent Sequence	Cluster Ra	ZR13_PvC1h	ZR13_PvC2h	ZR13_PvC4h	ZR13_PvC8h	ZR13_PvC12h	ZR13_PvC24h	ZR13_PvC48h	ZR13_PvC72h	ZR13_PvC96h
GGCACATGAATTCCCTTCGCTGACTGAATCACCAGCCTCT	99	0.76334055	-0.275539308	-0.148056847	0.159000906	0.526599143	0.994806105	-0.073548609	-0.808233785	0.309408195
CAACGTGTCAATGTTGAGTTTTTTCGTTGACTGATATTCC	100	0.471417615	0.000237413	-0.098115878	-0.316595354	0.584669491	-0.189112226	0.529538729	-0.043301353	-1.015322618
CACACGAACGCGTCTCATATCGAATATCGTTAACTAAACC	76	0.406751228	-0.152285772	-0.209268789	0.063269391	0.435079508	0.245704674	-0.070839261	0.33748632	0.287713199
CAACCAAACAGTTCCCACCATCAACTGTCGTTAACTGATC	53	0.357167755	0.011175912	-0.207789113	0.37634485	-0.127731266	0.202851048	-0.240445276	-0.054726889	-0.437434294
CAGGATAGTGAGTTAATAAACTCGTTGACTGACTCAGGCC	96	0.351402853	0.096281623	0.050296472	0.306995534	0.194631705	0.883648705	0.187851067	-0.075122805	-0.406092102
CAACGGTAGAGAATTCCAGATATTTCGTTAACTAAACGTC	97	0.303073882	-0.00062469	-0.441604292	0.432565111	0.073815249	0.879690054	-0.32779084	0.055410733	1.19287057
CAAACCGGACTTAGATAGTTTCCTAACTCGTTGACTGATC	83	0.295754351	0.201197565	-0.011088083	0.208330429	0.480027769	0.09986757	-0.085364576	-0.351626153	-0.744004762
GACGGGAAGTGCAATTGAACTACTATCGTTGACTAAACCC	40	0.251137995	0.272091918	0.000329078	0.30547701	0.0362188	-0.103922821	0.011376961	-0.263493465	-0.283134599
CACGGAACGGAAACGAGAGATTGTAATTGGTATACATGGG	30	0.214558914	0.06853248	0.127661649	0.260131485	-0.214795142	0.023418204	-0.027235145	0.109935352	0.064454814
GGCACACATAAGTTCATTATCAAGCTCGTTAACTAAAGAT	88	0.208196939	0.321620156	-0.020858244	0.365683449	-0.196333494	-0.036350396	0.132365058	0.034426648	0.059395725
CAAGTGTGGTTGAATATGTCCAAATCGTTGAGTGAACCCA	32	0.171689882	0.058804644	-0.089988833	0.569210515	-0.010338023	0.124529729	0.086089873	0.178469996	0.383757292
CAACAGTTAATAGCAAATTCAATCGTTAACTGAATGCCTC	89	0.170379255	0.558486502	-0.066886606	-0.038333698	0.862263992	-0.030581405	0.165364474	0.260752506	-0.776955494
GGGGGTGAGCAAGTCTCATAGGTCTACACGTTAACTGATA	59	0.13883029	-0.247845123	0.189441997	0.139884774	1.226325008	0.485188828	0.772936504	-0.3720903	-0.311159194
CAAGTAGTGATGTGTTGAGTCCCAATCGTTAACTGAGCTA	33	0.127997113	0.008385826	-0.257635127	0.012925278	-0.113819286	-0.12664305	0.356579762	0.22979431	-0.1413692
CATAGCAGGGAGACTTATACTCGTTGACTGATCGATATCC	95	0.107389039	-0.12523903	-0.334021891	-0.021783499	-0.004041347	0.20041484	-0.212406201	-0.236026322	0.126278834
CAAGCGCGTATCATGTCACCCTCAGATCTCGTTAACTGAT	27	0.067545937	0.304833554	0.429820275	0.312059838	0.008242984	0.00261833	0.17971271	-0.116228559	-0.220217252
CAGGGCCAGTTCACTTTCTATTAATCGTTAACTGATGTTC	85	0.038345766	-0.032819187	0.228339257	-0.098038729	0.12371772	0.578291572	-0.0685071	0.100075294	0.16153678
GGTAGAGGGTCAGAAAAGATAATTTCCTCGTTGACTGATC	73	0.031401094	0.20461693	0.600417579	0.387756292	-0.048312265	-0.039877063	0.264847537	0.970356428	0.340935645
CAAAACTACCAGTAGAAAGACCACATATCGTTGACTGACA	60	0.027198626	0.273600047	-0.174578188	0.199627438	-0.226405175	-0.193588219	0.221920528	0.075117293	-0.703229099
CAAAGGTGTGGAGGAATTTCAGTGAATCTATCGTTAACTG	90	0.021552726	-0.083324313	-0.006113071	0.279828927	0.5905779	0.206102345	0.495369317	0.280974474	0.468007382

Parent Sequence	Cluster Rai	R13_PvC1h	R13_PvC2h	R13_PvC4h	R13_PvC8h	R13_PvC12h	R13_PvC24h	R13_PvC48h	R13_PvC72h	R13_PvC96h
GGCACATGAATTCCCTTCGCTGACTGAATCACCAGCCTCT	99	3.104169164	0.802478003	1.474355811	1.361238832	2.16082822	2.506370356	1.263910662	1.033304777	2.601846366
CAACGTGTCAATGTTGAGTTTTTTCGTTGACTGATATTCC	100	2.812246229	1.078254724	1.52429678	0.885642573	2.218898568	1.322452025	1.866998001	1.798237209	1.277115553
CACACGAACGCGTCTCATATCGAATATCGTTAACTAAACC	76	2.747579842	0.925731539	1.413143869	1.265507317	2.069308585	1.757268925	1.266620011	2.179024881	2.58015137
CAACCAAACAGTTCCCACCATCAACTGTCGTTAACTGATC	53	2.697996369	1.089193223	1.414623545	1.578582777	1.506497811	1.714415299	1.097013996	1.786811673	1.855003877
CAGGATAGTGAGTTAATAAACTCGTTGACTGACTCAGGCC	96	2.692231467	1.174298934	1.67270913	1.50923346	1.828860782	2.395212956	1.525310339	1.766415756	1.886346069
CAACGGTAGAGAATTCCAGATATTTCGTTAACTAAACGTC	97	2.643902497	1.077392621	1.180808366	1.634803037	1.708044326	2.391254305	1.009668432	1.896949295	3.485308741
CAAACCGGACTTAGATAGTTTCCTAACTCGTTGACTGATC	83	2.636582965	1.279214876	1.611324575	1.410568355	2.114256847	1.611431822	1.252094696	1.489912408	1.548433409
GACGGGAAGTGCAATTGAACTACTATCGTTGACTAAACCC	40	2.591966609	1.350109229	1.622741736	1.507714936	1.670447877	1.40764143	1.348836233	1.578045097	2.009303572
CACGGAACGGAAACGAGAGATTGTAATTGGTATACATGGG	30	2.555387528	1.146549791	1.750074307	1.462369412	1.419433936	1.534982455	1.310224127	1.951473914	2.356892985
GGCACACATAAGTTCATTATCAAGCTCGTTAACTAAAGAT	88	2.549025553	1.399637467	1.601554414	1.567921375	1.437895583	1.475213855	1.469824329	1.87596521	2.351833896
CAAGTGTGGTTGAATATGTCCAAATCGTTGAGTGAACCCA	32	2.512518496	1.136821955	1.532423825	1.771448442	1.623891054	1.63609398	1.423549145	2.020008558	2.676195463
CAACAGTTAATAGCAAATTCAATCGTTAACTGAATGCCTC	89	2.511207869	1.636503813	1.555526052	1.163904228	2.49649307	1.480982847	1.502823746	2.102291068	1.515482677
GGGGGTGAGCAAGTCTCATAGGTCTACACGTTAACTGATA	59	2.479658904	0.830172189	1.811854655	1.342122701	2.860554085	1.996753079	2.110395776	1.469448262	1.981278977
CAAGTAGTGATGTGTTGAGTCCCAATCGTTAACTGAGCTA	33	2.468825728	1.086403137	1.364777531	1.215163204	1.520409791	1.384921202	1.694039034	2.071332872	2.151068971
CATAGCAGGGAGACTTATACTCGTTGACTGATCGATATCC	95	2.448217653	0.952778281	1.288390767	1.180454428	1.630187731	1.711979091	1.125053071	1.60551224	2.418717005
CAAGCGCGTATCATGTCACCCTCAGATCTCGTTAACTGAT	27	2.408374551	1.382850865	2.052232933	1.514297765	1.642472062	1.514182581	1.517171982	1.725310003	2.072220919
CAGGGCCAGTTCACTTTCTATTAATCGTTAACTGATGTTC	85	2.37917438	1.045198124	1.850751915	1.104199198	1.757946797	2.089855823	1.268952172	1.941613856	2.453974951
GGTAGAGGGTCAGAAAAGATAATTTCCTCGTTGACTGATC	73	2.372229708	1.282634241	2.222830237	1.589994219	1.585916812	1.471687189	1.602306809	2.81189499	2.633373817
CAAAACTACCAGTAGAAAGACCACATATCGTTGACTGACA	60	2.36802724	1.351617358	1.44783447	1.401865365	1.407823902	1.317976033	1.5593798	1.916655855	1.589209072
CAAAGGTGTGGAGGAATTTCAGTGAATCTATCGTTAACTG	90	2.362381341	0.994692998	1.616299587	1.482066853	2.224806977	1.717666596	1.832828589	2.122513036	2.760445553

Figure 19. A. Top 20 clusters ordered by highest Mean Cluster Enrichment Ratio Z-score at 1h. **B.** Non-standardized mean cluster enrichment ratios of the same clusters for comparison.

Figure 19 fails to highlight any outliers with significant deviation from the mean enrichment ratio, suggesting a narrow normal distribution of the data. These results now reflect the log-log sequence frequency plots from the time course library enrichment analysis, showing no statistically significant difference between the MHS and MHN incubation reactions. an overall lack of library selectivity towards MH susceptibility. Among the top 20 clusters, Figure 8 indicates that the top candidate at the 1h time point comparison is Cluster 99, with a Z-score of 0.76, suggesting it may hold sequences with preferential selectivity for MH susceptibility. This finding should be considered in the context of the cluster reflecting very low sequence frequency values, including its highest-abundance sequence, carrying a higher likelihood of introducing error and confounding the results. The top 3 clusters were plotted on log-log sequence frequency plots to ascertain whether any of their higher frequency sequences fall above the diagonal line, depicting preferential selectivity for MH susceptibility. The resulting plots are illustrated in Figure 9.

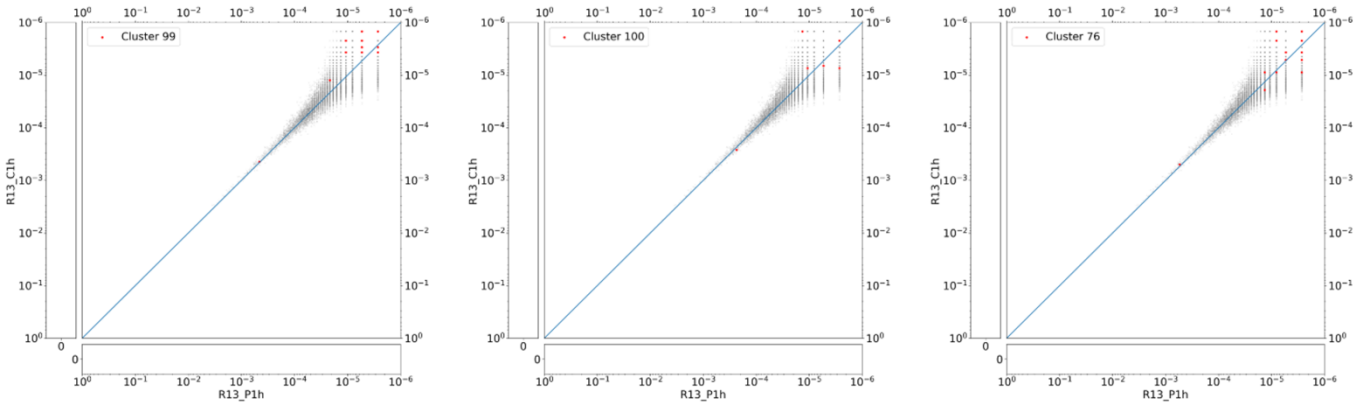


Figure 20. MH1 Selection – Log-Log Sequence Frequency Plots of Top 3 Sequence Clusters based on predicted higher selectivity for MH susceptibility at 1 hour, mapped against the total 1 hour time point population.

Across all three top clusters mapped in Figure 20, their most abundant sequences all lie on the diagonal line, indicating a lack of selectivity for MH susceptibility. The remaining sequences in each cluster gravitate into the very low abundance (low confidence) area of the plots, suggesting that the elevated Z-score is largely driven by the low-abundance sequences.

Unfortunately, despite attempts to optimize the sequencing data for high-accuracy comparisons, the malignant hyperthermia bioinformatic analysis has failed to identify any promising DNAzyme candidates with preferential selectivity towards MH susceptibility. The inability to identify a DNAzyme candidate may be a result of choosing a non-acute pathology, in which distinguishing biomarkers are in much lower concentration than common or non-specifically interacting serum components. Consequently, increased rounds of selection may be beneficial in continuing to tease out differences in the positive and counter selection serum pools. In the future, as it pertains to applying *in vitro* selection to a non-acute pathology, it may be especially beneficial to begin the selection protocol with a counter selection step rather than positive selection, in order to aggressively counter select any cross-reactive sequences in serum samples expected to be otherwise very similar.

4.3 Summary

In conclusion, this advanced bioinformatic analysis has acted as a diagnostic measure for evaluating the success of the two projects we have attempting involving *in vitro* selection in whole human serum. The results of the analysis have further substantiated the feasibility of performing *in vitro* selection in serum by virtue of confirmation of DNAzyme library enrichment over rounds of selection. Additionally, we successfully identified a two-part bioinformatic method for evaluating the progress and success of an *in vitro* selection protocol performed in a complex biological matrix, by first examining the enrichment of the library over time and in different media, followed by an in-depth examination of sequence clusters for identification of promising DNAzyme candidate sequences capable of detecting an intended target. While we recognize that the field of bioinformatics is incredibly robust and continuously expanding, our hope is that the analysis method provided in this chapter will constitute a springboard for further developments with cross-overs into deeper machine learning analysis. Such collaborations will undoubtedly improve the quality and depth of information extracted from recovered sequencing data, while providing guidance to fellow scientists pursuing *in vitro* selection in complex biologic matrices.

4.4 Experiments

4.4.1. Processing of NGS sequencing data

Purified PCR products were sequenced using paired-end Next-Generation Sequencing (NGS) using an Illumina Miseq system at the Farncombe Metagenomics Facility, McMaster University. Raw sequencing reads were first trimmed of their primers using Geneious. The resulting 40 nt reads were filtered for quality using PrinSeq v0.20.4 to make sure only high-quality reads were used for further analysis. All sequences containing any bases of Phred scores < 20 (base-call probability < 99%) were eliminated. Sequence alignment and dereplication to identify unique sequences was performed and their copy numbers were extracted. Using a clustering algorithm CD-HIT-EST, sequences were grouped into clusters. 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. Enrichment ratios and unique sequence frequency values were calculated using the equations profiled in Chapter 4 Results.

4.4.2. Log-Log Sequence Frequency Plots

Using unique sequence frequencies, Log-Log Sequence Frequency plots were created using Matplotlib version 3.7.1. Each sequence's coordinates corresponded to the unique sequence frequency in each of the two media/time points/conditions being compared.

4.4.3. Cluster Analysis

The clustering algorithm mentioned above was used to identify sequence clusters with 90% identity conservation of the random domain. Cumulative cluster frequencies in each medium/time

point/condition were calculated by taking the mean of all sequence frequencies in the cluster. Mean Cluster Enrichment ratios were obtained by calculating the ratios of the cumulative cluster frequencies in the two conditions compared. Cluster kinetics analysis was performed using GraphPad Prism 10.0.3 software. Each cluster was modelled by non-linear regression using the one-phase association equation $Y = Ymax [1-e^{-kt}]$, employing the mean frequency of each cluster across the nine time points submitted for sequencing to fit the pseudo-first order model. The constraints imposed were $Y_0 = 0$, and Plateau < 100. The plateau (Bmax) of each predicted model reflects the frequency of each cluster in the total Round 14 population. The relevance of this caveat is to differentiate the plateau values in this predicted model from the values we expect to see in a pure experiment with individual sequences. Based on the modelled data, rate constant K_d values are generated.

4.4.4 Normalization of mean enrichment ratios

To account for the read depth discrepancies, the cluster mean enrichment ratios in positive and counter selection serum incubations were normalized against the mean population enrichment ratio in each serum.[5] Specifically, we calculated the Z-score of each cluster's mean enrichment ratio by subtracting it from the mean population enrichment ratio.[6] In doing so, we obtained the degree of deviation of each cluster from the mean enrichment ratio of the population, normalizing the data against a mean value of 0.[6]

4.5 References

1. Smith MD, Case MA, Makowski EK, Tessier PM. Position-Specific Enrichment Ratio Matrix scores predict antibody variant properties from deep sequencing data. *Bioinformatics*. 2023;39(9). doi:10.1093/bioinformatics/btad446
2. Komarova N, Barkova D, Kuznetsov A. Implementation of High-Throughput Sequencing (HTS) in Aptamer Selection Technology. *Int J Mol Sci*. 2020;21(22):8774. doi:10.3390/ijms21228774
3. Weiss Z, DasGupta S. REVERSE: a user-friendly web server for analyzing next-generation sequencing data from *in vitro* selection/evolution experiments. *Nucleic Acids Res*. 2022;50(W1):W639-W650. doi:10.1093/nar/gkac508
4. Fowler DM, Araya CL, Gerard W, Fields S. Enrich: software for analysis of protein function by enrichment and depletion of variants. *Bioinformatics*. 2011;27(24):3430-3431. doi:10.1093/bioinformatics/btr577
5. Boeva V, Popova T, Lienard M, et al. Multi-factor data normalization enables the detection of copy number aberrations in amplicon sequencing data. *Bioinformatics*. 2014;30(24):3443-3450. doi:10.1093/bioinformatics/btu436
6. Dillies MA, Rau A, Aubert J, et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform*. 2013;14(6):671-683. doi:10.1093/bib/bbs046

Chapter 5. *in vitro* selection in human serum

5.1 Introduction

The prior three chapters have successfully proven the feasibility of performing *in vitro* selection in a complex biological matrix such as whole human serum. Specifically, we have demonstrated that we can apply selective pressures to a DNAzyme library in serum and observe correlated enrichment in the library towards a target of interest. However, as with any proof-of-concept study, pitfalls and areas of improvement have been identified, and will serve the basis of discussion in this chapter.

Undoubtedly, *in vitro* selection is a powerful tool for the targeted and systematic pursuit of biomarkers and their detection probes. Coupled with the advent and booming development of bioinformatic tools, the future of *in vitro* selection research is bound to be fruitful and promising. The true strength of *in vitro* selection rests in its ability to identify a biomarker without *a priori* knowledge of the target pursued. However, this strength can also present as a challenge in a complex biological matrix such as human serum, due to the difficulty in tracking meaningful target-mediated DNAzyme cleavage while accounting for non-specific interactions with serum components or degradation by nucleases. Additionally, challenged with seroconversion have shed light on the impact of sample preparation on the success of *in vitro* selection.

Consequently, by identifying challenges observed over the course of the two selections performed and exploring their impacts on the effectiveness of the selection, we hope to highlight solutions to optimize the *in vitro* selection approach in whole human serum, towards the development of a standardized methodology.

5.2 Results

5.2.1 Enrichment comparison of pancreatic cancer and malignant hyperthermia *in vitro* selections

Based on the challenges observed with counter selection in the pancreatic cancer project, the malignant hyperthermia protocol was amended to allow for significantly more aggressive counter selection by increasing the incubation time with counter selection serum from 2 hours to 16-24 hours. The positive selection incubation time was also increased to 4 hours to avoid paring down the DNAzyme library excessively from the onset of the experiment. These changes subsequently prompted an inquiry into the effect of the increased incubation times on the efficacy of the selection. To monitor the efficacy, we opted to use the enrichment ratio and compare the pancreatic cancer and malignant hyperthermia selections' enrichment ratios round-by-round. The results of this comparison are illustrated in Figure 1.

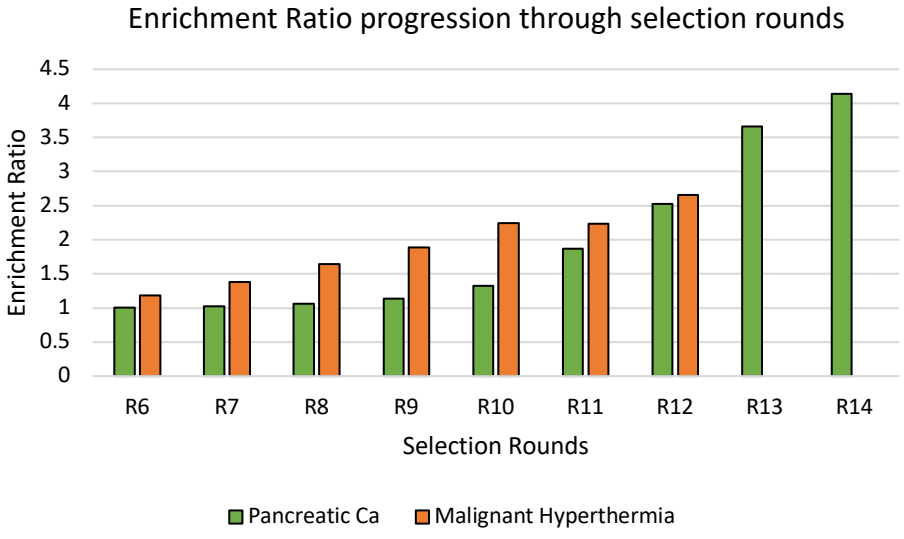


Figure 1. Comparison of DNAzyme libraries' enrichment over course of *in vitro* selection protocol in pancreatic cancer and malignant hyperthermia.

A readily apparent observation from Figure 1 is the malignant hyperthermia selection's higher enrichment in earlier rounds as compared to the pancreatic cancer selection, suggesting that longer incubation times for both positive and counter selections may maximize round-to-round enrichment. This is likely attributable to the longer time intervals allowing more extensive interactions between the DNAzyme library and serum components in each round of selection. The discrepancy in enrichment appears to taper off at Round 12, when the pancreatic cancer counter selection incubation time was increased to 24 hours – the same amount of time as in the malignant hyperthermia protocol. Interestingly, with increasing counter selection incubation times (up to 48h at Round 13 and 14), the pancreatic cancer enrichment ratio begins to rise significantly.

Therefore, our conclusion from this comparison is that *in vitro* selection efficiency appears to benefit from prolonged incubation time with serum, as evidenced by longer incubation times for both counter and positive selection appearing to improve round-to-round enrichment of DNAzyme libraries undergoing *in vitro* selection in human serum. This effect is likely attributable to ample time required by the DNAzyme library to interact with the myriad of serum components, and more importantly with the target of interest, particularly when the identity and concentration of the target is unknown and presumably low. Whether increasing the incubation time of positive and counter selection translates into shorter *in vitro* selection protocols by virtue of fewer rounds being necessary to reach adequate enrichment is difficult to ascertain without additional research.

The implications of a shorter selection protocol may include limiting the detrimental effects of non-specific serum interactions and/or DNAzyme degradation by limiting the number of selection rounds and in doing so limiting the number of opportunities for such interactions. Furthermore, in

a complex and dynamic matrix like human serum, decreasing the number of selection rounds performed may offer more control over the intended counter selection effect, particularly when positive and counter selection samples are so closely similar. In this instance, a sequence that should cleave in counter selection serum but does not, is given another opportunity once carried over into positive selection serum. A cleavage mistake in positive selection serum is subsequently amplified, and becomes significantly more difficult to correct in subsequent rounds.

A second criterion of comparison between these two selections was alluded to in Chapter 3, namely the type of pathology targeted. For this thesis exploration, we selected an acute pathology and a non-acute pathology to investigate using *in vitro* selection experiments. In addition to examining the scope of DNAzyme detection through diversification of pathologies investigated, concerns were raised during the pancreatic cancer selection with regards to acute complex pathologies like cancer introducing an overwhelming amount of complexity to the biomarker landscape from upregulated immune and oncogenic activity, and confounding the results. However, the exploration of *in vitro* selection in a stable non-acute pathology has not demonstrated to be more fruitful despite increased counter selection incubation. Based on the results and performances of the two selection protocols undertaken, it appears that *in vitro* selection may prove more effective when applied active pathologies. Despite the metabolic complexity described accompanying an active disease process, active pathologies are also more likely to upregulate biomarkers specific to the pathology investigated, or upregulate a common target with concentration parameters specific to the pathology investigated. Either of these features would facilitate *in vitro* selection in identifying differences between the positive and counter selection serums, therefore enhancing the success of the selection.

5.2.2 The impact of blood sample collection and serum preparation on *in vitro* selection

Throughout both selections, a common challenge identified was the difficulty in reconciling DNAzyme differences in cleavage activity based on method of serum conversion and age of collected samples. We attempted to circumvent this challenge by performing a DNAzyme cleavage assay prior to the start of the first selection, in order to pre-emptively identify whether the method of serum preparation or blood fraction leads to aberrant cleavage patterns, and identify the optimal blood fraction and preparation method for conducting the selection experiment. The results of this DNAzyme cleavage assay are displayed in Figure 2.

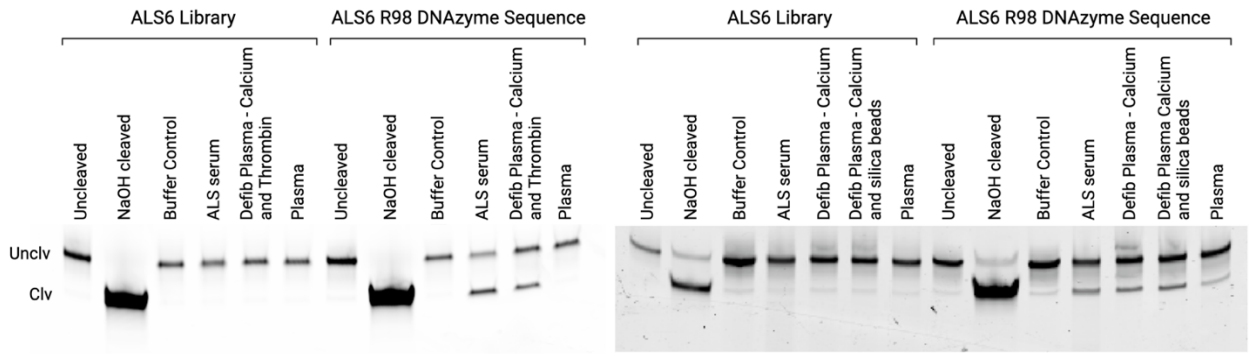


Figure 2. DNAzyme cleavage assay for ALS6-R98 DNAzyme sequence and ALS6 naïve library.

The DNAzyme cleavage assay depicted in Figure 2 was designed to test a DNAzyme sequence previously identified by our research group for detection of amyotrophic lateral sclerosis (ALS). The DNAzyme sequence was tested against its naïve selection library for comparison, in multiple blood fractions and using multiple serum conversion methods. The biological matrices explored include plasma, whole-blood derived serum, and defibrinated plasma obtained by calcium

addition, calcium and silica beads addition, and thrombin and calcium addition. The results of this assay demonstrated equivalent cleavage in whole-blood derived serum and defibrinated plasma across all preparation methods, and considerably less cleavage in plasma. Importantly, this assay did not show differences in cleavage among defibrination methods. Consequently, we felt comfortable proceeding with the plasma samples obtained from the Ontario Tumor Bank, and opted for the defibrination method that introduced the least amount of change to the samples – addition of calcium +/- silica beads. In doing so, we avoided introducing thrombin to the selection reaction as a potential target for the DNAzyme library to erroneously interact with.

Despite this experiment, notable differences were observed in the cleavage performance of DNAzyme candidates in both selections, particularly in the sensitivity and specificity testing, suggesting that the results of the assay performed above may be specific to the sequence tested, and serum preparation methods are not equivalent. Beyond the performance of the DNAzymes, differences in DNAzyme degradation patterns were also noted across different serum preparation methods, and across both selections. While a distinct pattern and theory to unify all results observed cannot be discerned, an enumeration and discussion of the cleavage and degradation patterns observed will be detailed below.

Beginning with the pancreatic cancer project, four distinct categories of serum samples will be explored, reflecting the variety of serum samples employed for the selection and sensitivity and specificity analysis, and categorized based on criteria of age of collected samples, method of serum derivation, and target. Based on age of collection, samples can be broken down into the banked OTB samples and the fresh collected samples from local area patients. Based on serum derivation,

samples can be broken down into plasma-derived or whole-blood derived serum. All OTB samples were converted to serum from plasma through defibrination by addition of calcium chloride. Among the fresh collected samples, one blood sample was collected in an EDTA-coated vacutainer tube and consequently processed for plasma, followed by defibrination to serum, as per the OTB serum preparation protocol. The remaining fresh collected samples were converted directly to serum from whole blood by allowing the blood sample to clot and centrifuging the clotted sample. Lastly, the OTB samples targets have been previously described (pancreatic, breast, colorectal, lung, prostate), while the fresh collected samples include 5 pancreatic cancer samples and 5 healthy participant samples. During the sensitivity and specificity analysis, consistent cleavage and DNAzyme degradation patterns were noted across multiple categories of samples. A schematic breakdown of the samples and their performance when tested with the 14-2 DNAzyme is provided in Figure 3.

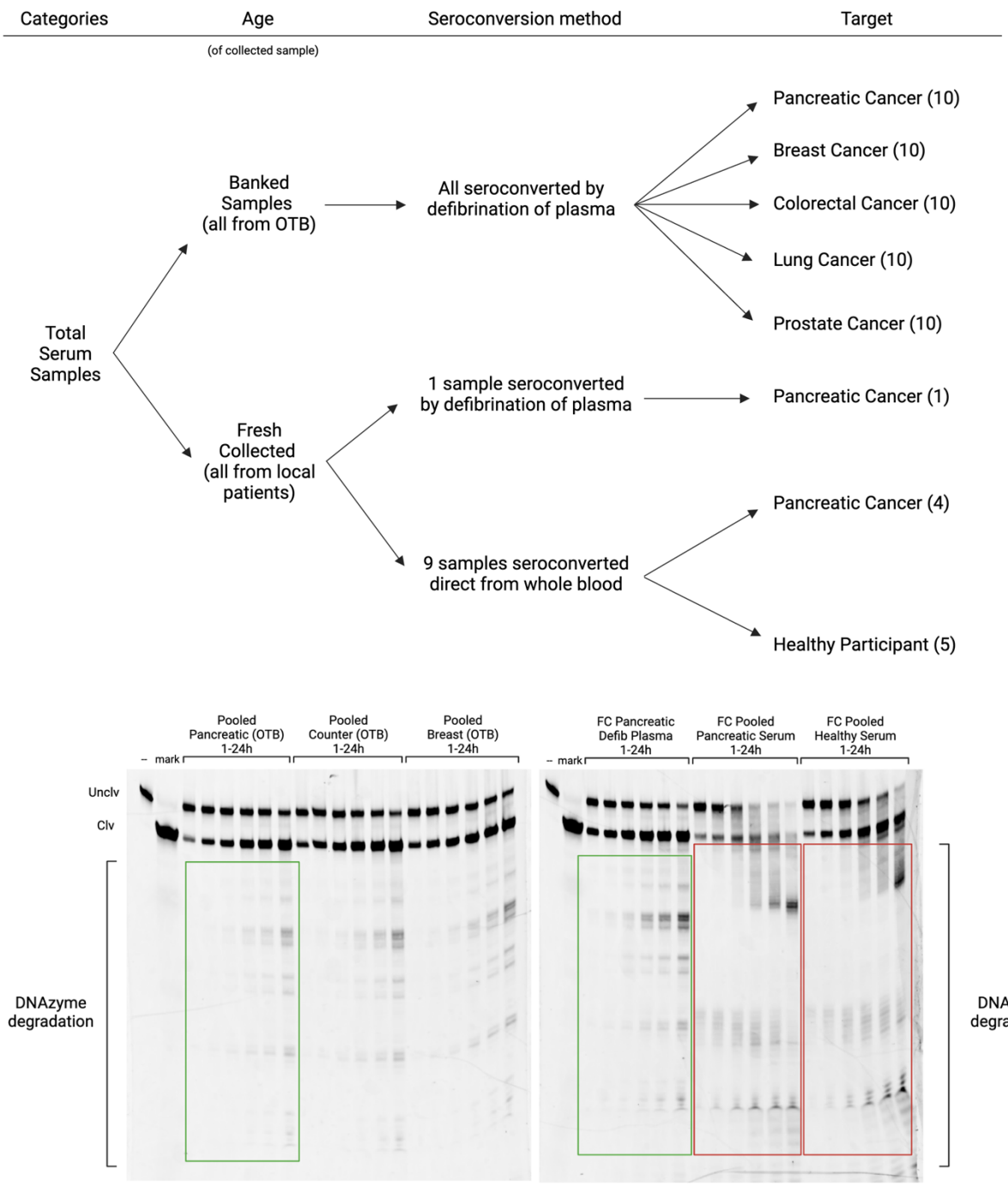


Figure 3. A. Pancreatic cancer project serum sample breakdown by numerous categories. **B.** Gel depiction of DNAzyme 14-2 cleavage and degradation patterns across 24h time courses.

Figure 3 indicates consistent cleavage patterns of the 14-2 DNAzyme across all samples and time points. This expected finding is demonstrated by the presence of uncleaved and cleaved bands at their expected positions on each gel relative to the cleavage markers. However, there appear to be distinct nuclease-mediated DNAzyme degradation patterns across the samples. For instance, all plasma-derived OTB samples show minimal degradation of the uncleaved and cleaved DNAzyme bands, as evidenced by a gradual decrease in fluorescence density of the uncleaved band with concomitant increase in the cleavage band over time, and minimal downstream smearing. Furthermore, the degradation patterns downstream from the cleavage band demonstrate a consistent pattern across all OTB samples, regardless of target. This finding suggests the degradation observed can be attributed to nuclease activity common yet specific to all cancers, or to non-specific nuclease degradation. Remarkably, the fresh collected defibrinated plasma pancreatic cancer sample follows the same cleavage and degradation patterns as the OTB samples, with marginally increased downstream degradation of the 14-2 DNAzyme, but yielding the same fragment patterns. This suggests that provided the samples are processed in the same manner and adequately stored, the collection age of the samples bears little effect on their quality over time, since the observed DNAzyme activity remains near identical in the new and banked samples. This conclusion reinforces the generalizability of banked OTB samples to fresh collected samples and alleviates concerns of significant loss of biodiversity in the samples over time.

While sample age does not appear to affect DNAzyme activity and degradation patterns, the same cannot be said for seroconversion method. Despite concomitant collection, the whole-blood derived serum samples differ greatly from the fresh collected defibrinated plasma serum sample. Specifically, the whole-blood derived samples appear to enable considerably more nuclease

breakdown of the DNAzyme, as evidenced by notable fragment smearing on the gel downstream from the cleavage band, along with significant degradation of the uncleaved DNAzyme band over time. Interestingly, the pancreatic cancer and healthy participant serum samples yield similar degradation patterns downstream from the cleavage band, despite belonging to different target categories, suggesting that the extent of degradation and its pattern is reliant on the seroconversion method employed and is consequently higher in serum derived directly from whole blood. Of note, the degradation effect is more pronounced in the fresh collected pancreatic cancer pooled serum, with extension of DNAzyme degradation to the cleaved band as well. Overall, the extensive degradation observed in this seroconversion method may suggest higher preservation or concentration of nucleases in serum converted directly from whole blood, and possibly upregulated nuclease activity in cancer samples.

An explanation for the seroconversion method effect on nuclease activity may be offered by the lack of chelating agents employed in obtaining serum from whole blood, unlike the plasma defibrination method of seroconversion. Since nucleases also depend on divalent metal ions[1,2] to perform their catalytic function of nucleic acid degradation, the presence of EDTA in the plasma-derived samples (from blood collection tubes) may be impairing their activity such that less DNAzyme degradation is observed in the defibrinated plasma samples. With regards to the seemingly upregulated nuclease activity in the pancreatic cancer whole-blood derived samples, this finding is in keeping with recent literature in the field of cancer metabolomics, which indicate that nucleases are not only found to be upregulated in cancer, but can in fact serve as biomarkers of cancer diagnosis.[3]

Ultimately, the conclusion from the pancreatic cancer project suggests the most significant determinant of DNAzyme probe degradation in serum is the method of seroconversion, defying sample age and even target category. This finding is particularly substantiated by the fresh collected defibrinated sample, which behaves near-identically to the OTB samples despite having a different source of collection and not being involved in the *in vitro* selection protocol.

We subsequently turned out attention to the malignant hyperthermia selection to determine if the seroconversion methods employed had effects of DNAzyme cleavage and degradation. All ninety MHS/MHN samples employed in the sensitivity and specificity analysis, including the twenty samples employed in the *in vitro* selection protocol came from the same source – the MHIU. Although late in the project and halfway through the sensitivity and specificity analysis, an additional whole-blood derived serum sample was obtained locally from a patient recruited into our study with a medical history of two separate malignant hyperthermia reactions (but no CHCT diagnosis). This sample was not included in the sensitivity and specificity analysis due to its late arrival and lack of gold standard validation, but was independently tested with the candidate DNAzymes identified in Chapter 3.

Despite coming from the same source, the MHIU samples did not prove equally susceptible to seroconversion with calcium chloride addition (+/- silica beads). Specifically, the first twenty samples received for initiation of the *in vitro* selection protocol were successfully seroconverted with 1-2 doses (12.5-25mM) of calcium chloride. However, the subsequent seventy samples received for the sensitivity and specificity analysis could not be forced over the clotting threshold regardless of the number of calcium doses added (up to six). Given that the DNAzyme selection

was already completed, a decision was made to convert the new MHIU samples with thrombin. Consequently, all seventy plasma samples seroconverted within seconds or minutes of thrombin addition, without any added calcium. This prompted the question of whether heparin-coated tubes were used for collection of these samples, given heparin's neutralization of thrombin via activation of antithrombin[4], in addition to an apparent lack of calcium chelation based on successful seroconversion with thrombin alone. However, we were reassured the samples were collected in EDTA-coated vacutainer tubes. The collection age of the samples was considered, possibly correlating with a higher concentration of EDTA in plasma; however, given the ease with which the fresh collected defibrinated sample in the pancreatic cancer project was seroconverted with calcium chloride, this theory was deemed once again unlikely. Ultimately, the causative agent responsible for the inability to seroconvert the testing samples with calcium chloride remains unclear. However, this challenge has shed light on the reliability of plasma defibrination as a seroconversion method.

Further curious observations were noted with the malignant hyperthermia serum samples particularly in the sensitivity and specificity stage of the project. As previously alluded to in Chapter 3, the new serum samples received for candidate DNAzyme testing showed marginal cleavage with all candidate DNAzymes regardless of positive or negative sample status. A possible explanation proposed was overfitting of the DNAzyme library to the samples employed in the selection. Barring significant changes in the metabolomic landscape of the selection and testing serum samples, this theory remains unlikely. A second possibility for this observation is the chelation effect of leftover EDTA in the serum samples recovered following plasma defibrination with thrombin. Unlike the selection samples which were converted by overwhelming and

saturating the plasma EDTA (introduced by the blood sample collection tube), the testing samples did not receive additional calcium. Therefore, while clotting was achieved without calcium addition, it is possible the calcium-chelating effect of EDTA in solution is prohibiting DNAzyme folding and subsequent target interaction and cleavage. We attempted to substantiate this theory by assessing whether a lack of nuclease activity is concurrently observed in samples with marginal DNAzyme activity. A thorough comparison of DNAzyme cleavage and degradation patterns of all individual patient samples upheld this theory, with a demonstrated lack of observed nuclease-mediated DNAzyme degradation in most serum samples exhibiting minimal DNAzyme activity. However, this analysis also led to an observation of contrasting degradation patterns across serum samples, seemingly correlated with their collection age. Figure 4 depicts notable gel patterns emerging from the sensitivity and specificity analysis of the 12-3 DNAzyme, along with a gel depicting the 14-2 DNAzyme in fresh collected serum samples, for comparison.

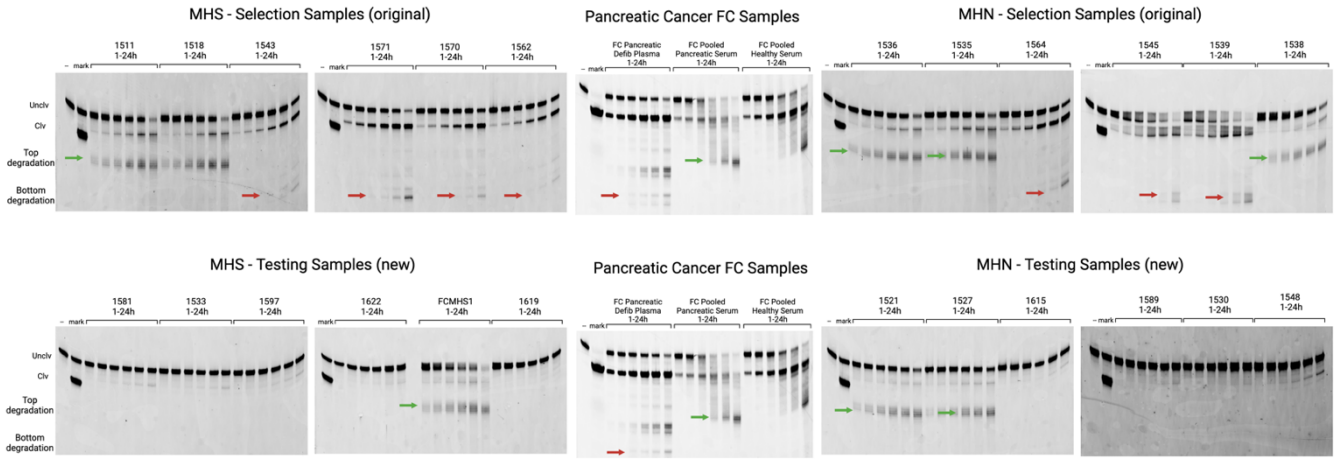


Figure 4. Comparison of cleavage and degradation patterns of DNAzyme 12-3 in selection and testing samples. Pancreatic Cancer fresh collected samples with DNAzyme 14-2 included for DNAzyme degradation comparison. FCMHS1 = Fresh Collected Malignant Hyperthermia Sensitive (local patient serum sample).

Figure 4 indicates two different patterns of degradation are present in the malignant hyperthermia defibrinated plasma samples, nicknamed “top band degradation” (green arrow) and “bottom band degradation” (red arrow). This effect was not observed in the pancreatic cancer defibrinated samples, all of which appeared to uniformly demonstrate the pattern of “bottom band degradation”, including the fresh collected sample. The pancreatic cancer fresh collected samples were included in Figure 4 to aid in additional pattern recognition. Based on the pancreatic cancer samples, “top band degradation” appears to match the whole blood-derived serum sample degradation, while the “bottom band degradation” appears to match defibrinated plasma sample degradation. This effect is extended to the fresh collected pancreatic cancer defibrinated plasma sample. Unsurprisingly, all malignant hyperthermia selection samples (samples used to perform the *in vitro* selection protocol) show degradation of the 12-3 DNAzyme probe. This finding is expected and consistent with their calcium-only seroconversion method, leading to an abundance of calcium present in the serum sample. However, their degradation patterns curiously appear to diverge at sample 1538 and 1539, such that samples numbered 1538 and under all demonstrate “top band degradation” consistent with blood-derived serum degradation. Meanwhile, samples numbered 1539 onwards appear to demonstrate “bottom band degradation” consistent with defibrinated plasma degradation patterns. Interestingly, this observation stands regardless of MHS or MHN status. These findings may be explained by the collection age of the samples, or by different practices adopted over time in the collection and processing methods of the blood samples.

Additionally, two new testing samples (1521 and 1527) are illustrated in Figure 4. These samples could not be seroconverted with calcium chloride, but were successfully seroconverted with thrombin. Yet, they still appear to demonstrate DNAzyme cleavage and nuclease degradation

patterns. These patterns are specifically in keeping with their chronological ID (<1539) – namely "top band degradation", further substantiating the observation made regarding the degradation patterns of the original serum samples used for the selection. While both DNAzyme and nuclease degradation appear to be milder in new samples 1521 and 1527 in comparison to the original serum samples received for conducting the selection experiment, these activities stand out among the remaining testing samples which demonstrate minimal DNAzyme cleavage and virtually no nuclease degradation. This observation once again brings into question the calcium concentration in the plasma of these two samples, given the inability to seroconvert them by oversaturating the EDTA with calcium chloride, yet noting persistence of calcium-dependent DNAzyme and nuclease activity in the thrombin-only converted serum of these samples. Evidently, the unknown yet incongruent concentrations of calcium in the serum samples and their effects on DNAzyme function are recurring concerns in the malignant hyperthermia project.

Consequently, an experiment was designed to assess the effect of calcium concentration on DNAzyme function, and identify whether the seroconversion method employed could be responsible for the discrepancies noted in DNAzyme activity between the original and new serum samples. Specifically, this experiment aimed to answer the following questions:

1. Is there an optimum calcium concentration for DNAzyme activity in serum?
2. Can DNAzyme activity be restored in thrombin-converted serum samples with addition of calcium to the recovered serum?
3. Does DNAzyme activity also decrease in original patient samples when seroconverted with thrombin, and is the activity subsequently restored with addition of calcium to the serum?

In order to perform this experiment, eight patient samples were employed – four original MHS samples (1518, 1543, 1570, 1571), and four MHS samples from the new batch of testing samples (1614, 1654, 1655, 1664). The latter four samples were among the only ones successfully seroconverted with calcium chloride during the initial calcium chloride seroconversion test of the new testing samples. Their serum was recovered and stored, and subsequently used for this experiment. Consequently, the eight samples identified underwent seroconversion using the two methods described - once with calcium chloride and once with thrombin, effectively yielding two distinctly obtained serums per patient sample. These two serums were fractionated and incremental doses of calcium were added to each fraction. This process was repeated for each patient sample. The calcium concentrations examined were designed to reflect the incremental doses of calcium administered in the calcium chloride seroconversion method. In this method, each dose resulted in a 12.5mM increase in the serum calcium concentration. Consequently, samples converted with one dose yielded a calcium concentration of 12.5mM over the baseline serum calcium concentration, while samples requiring two or three doses yielded concentrations of 25mM and 37.5mM, respectively over the baseline serum calcium concentration. We opted to explore the effect of incremental calcium doses until 100mM above the baseline. Once all serum fractions were created, each fraction was incubated with the 12-1 DNAzyme and 12-hour time courses were performed. The 12-1 DNAzyme was employed in this experiment given its best cleavage performance in the sensitivity and specificity analysis. Control reactions with selection buffer (no serum) were included. The cleavage percentage results were subjected to kinetic analysis using a pseudo-first order reaction model to identify any rate constant trends and concentration correlations. The results of this experiment are depicted in Table 1.

Condition	1518	1543	1570	1571	1614	1654	1655	1664	SB
Ca 0	-	-	-	-	-	-	-	-	0.000866
Ca 12.5	0.001016	0.002661	0.004064	0.005533	-	-	-	-	0.001217
Ca 25	0.00252	0.01286	0.001521	0.002956	-	-	0.003004	0.003004	0.001353
Ca 37.5	0.002284	0.001635	0.001593	0.001248	0.003185	0.007317	0.003045	0.002039	0.002349
Ca 50	0.002148	0.001558	0.001393	0.0012	0.001809	0.002336	0.001866	0.001577	0.001259
Ca 62.5	0.001686	0.000507	0.000659	0.000756	0.001641	0.001992	0.001698	0.001126	0.001874
Ca 75	0.001425	0.000892	0.000859	0.000871	0.00102	0.000942	0.000827	0.000791	0.003183
Ca 87.5	0.001566	0.000961	0.00099	0.001583	0.000962	0.000818	0.000686	0.00107	0.002757
Ca 100	0.001617	0.001137	0.001185	0.001717	0.000856	0.000961	0.000773	0.00096	0.001043
Thr 0	0.001205	0.000591	0.0021	0.000631	0.000655	0.000863	0.001915	0.00078	0.000942
Thr 12.5	0.004711	0.004908	0.004986	0.006062	0.003824	0.005993	0.009107	0.00835	0.000777
Thr 25	0.008128	0.01459	0.001107	0.006875	0.00178	0.002504	0.00716	0.000918	0.000785
Thr 37.5	0.006228	0.004052	0.001195	0.001711	0.000512	0.000582	0.001647	0.000782	0.00077
Thr 50	0.002356	0.001405	0.00141	0.000963	0.000406	0.001049	0.001808	0.002313	0.000826
Thr 62.5	0.001956	0.001352	0.001585	0.000786	0.000464	0.001	0.000894	0.001729	0.000875
Thr 75	0.00185	0.001362	0.000591	0.000988	0.000532	0.000886	0.000856	0.003063	0.001161
Thr 87.5	0.001741	0.001135	0.000583	0.001222	0.000199	0.000816	0.000989	0.001884	0.001345
Thr 100	0.00085	0.001352	0.000875	0.000546	0.000682	0.001082	0.001002	0.000573	0.002019

Table 1. Results of calcium concentration and seroconversion experiment, expressed as the values of the rate constant of each reaction. SB = Selection Buffer. Conditional formatting performed for each column to highlight the fastest rate constants and their correlated calcium concentrations for each patient sample. Plateau = 100%.

Given the shorter timeline of the time courses, the accuracy of the rate constant values is weak; however, qualitative trends can still be observed and contextualized with cleavage percentage data. Across all samples seroconverted with thrombin, it appears that addition of 1-2 doses of calcium chloride restores/increases DNAzyme activity. This conclusion is especially apparent in the raw cleavage data, whereby near identical cleavage percentages are observed when comparing the 1-2 dose calcium-seroconverted samples and 1-2 dose thrombin-seroconverted samples. This effect was present in every patient sample, and is further substantiated by the two highest rate constants in each patient sample corresponding to the mirror serum fractions (same calcium concentration

regardless of seroconversion method). Interestingly, when comparing absolute cleavage values, the raw cleavage percentages across all samples tested in this experiment remained considerably lower than the original samples tested in the sensitivity analysis. This is likely attributed to the ample time required for preparation of each serum fraction and subsequent preparation of incubation reactions, leading to potential target degradation while the serum samples were out at room temperature. Nevertheless, a 6-30x improvement in cleavage percentage was observed with calcium correction when comparing 1-2 dose thrombin-converted samples to no calcium addition.

A second observation made in this experiment relates to optimal calcium concentration in serum for peak DNAzyme function. While we have demonstrated that serum samples require calcium for DNAzyme activity, this effect is limited to 1-3 doses of calcium. Beyond 3 doses yielded a gradual decline in DNAzyme activity, with raw cleavage percentages similar to no addition of calcium. Therefore, the optimum calcium concentration for peak DNAzyme function appears to mirror the amount of calcium added to seroconvert the serum samples employed in the selection. This conclusion begs the question of whether we are selecting DNAzymes for calcium detection rather than disease-specific biological target detection. Fortunately, the control reactions with selection buffer confirmed this is not the case. While DNAzyme cleavage in buffer mildly increased with calcium concentration, across all buffer-only reactions the peak DNAzyme cleavage values remained below 1% through 12 hours, consistent with the sensitivity and specificity analyses in both selections. Consequently, while calcium does not appear to be the target responsible for mediating DNAzyme cleavage, it does appear that the DNAzymes selected require artificially heightened calcium concentrations in serum samples obtained from plasma defibrination in order to function optimally. This theory may also offer insights into understanding the observed

DNAzyme activity discrepancies in whole-blood derived serum compared to defibrinated plasma. Furthermore, the presence of additional thrombin in serum does not appear to affect DNAzyme cleavage, as demonstrated by near identical cleavage percentages in mirror calcium/thrombin seroconversion serum fractions (same calcium concentration, different seroconversion method), and no effect on DNAzyme cleavage in the buffer only reactions.

Notably, nuclease activity in this experiment was also restored with calcium addition. Additionally, thrombin-converted serum samples demonstrated comparatively lower nuclease activity across all calcium concentrations compared to their calcium-converted counterparts, across all patient samples despite restoration of DNAzyme activity. Whether this finding points to thrombin exhibiting a protective effect over DNAzyme degradation is unclear and requires further study.

Overall, these results indicate that seroconversion method does affect DNAzyme function and activity, particularly as it relates to addition of calcium to the serum sample, and regardless of magnesium ions being added to each selection reaction through selection buffer. Given the systematic approach of the *in vitro* selection protocol, perhaps it should come as no surprise that the conditions of the selection need to be perfectly matched in order to emulate DNAzyme activity. Furthermore, the challenges presented in the malignant hyperthermia experiment combined with the findings of this calcium concentration experiment substantiate the importance of a standardized *in vitro* selection protocol being developed with the fewest number of alterations to the biological samples studied, so as to decrease the chance of introducing error or confounding into the experiment, along with unforeseen challenges manipulating complex biological samples.

In conclusion, while some of the puzzling observations regarding sample preparation and DNAzyme activity in these two selection experiments were explained through the calcium concentration experiment, no unifying theory appears to account for all the discrepancies observed with seroconversion and DNAzyme degradation patterns. Based on the results of these two selections, it appears the optimal seroconversion method is whole blood-derivation of serum. This method eliminates the two-step approach of plasma separation followed by defibrination, which has been proven to be unreliable in an experiment that requires strict adherence to selection conditions and reagent concentrations. Furthermore, whole-blood derivation of serum avoids unnecessary altering of the biological landscape of native serum with addition of chelating (or other anticoagulant) agents, followed by saturation of the agent with calcium. The addition of thrombin further distorts the metabolic composition of the sample. The plasma defibrination protocol ultimately translates to less control over the concentration of calcium in selection reactions, which is of particular concern given the dependence of DNAzymes on divalent metal ions for proper folding and function. Evidently, nuclease activity remains a challenge for whole-blood derived serum, given its correlation with higher nuclease degradation of DNAzyme sequences. Consequently, the next section will address strategies to mitigate nuclease degradation effects on *in vitro* selection experiments.

5.2.3 Non-specific serum interactions negatively affect *in vitro* selection

In addition to serum preparation methods, another challenge regularly alluded to over the course of the selections performed is the difficulty in discerning target-mediated DNAzyme cleavage from non-specific interactions with common serum components. This phenomenon invariably affects the ability of counter and positive selection to effectively separate a DNAzyme library based on interactions with targets of interest. Unfortunately, this phenomenon reflects the reality of attempting selection in serum such that a certain amount of non-specific serum component interactions are unavoidable without altering the biological landscape of the complex biological matrix explored. Ideally, the effect of these interactions would be limited over time by frequent and aggressive counter selection, in favour of filtering out sequences undergoing meaningful target-mediated interactions specific to a pathology of interest.

The interference of non-specific interactions with the selection protocols has been most notably observed through nuclease-driven DNAzyme degradation. Nuclease-mediated degradation has been acknowledged in gel, kinetic, and sequencing analyses. Specifically, all three analysis categories demonstrate this phenomenon when examining DNAzyme time course reactions. In an ideal reaction scenario, any DNAzyme cleavage observed over time is attributed to meaningful target interactions, leading to an increase in fluorescence generation that is dependent on the concentration of DNAzyme in solution. This effect is based on the assumption that the concentration of DNAzyme in the selection reaction greatly exceeds any potential target concentration. Consequently, this ideal reaction is therefore best approximated by the pseudo-first order kinetic model, and depicts an enzymatic reaction with an eventual plateau which remains stable over time once all DNAzymes in the reaction have undergone target-mediated cleavage.

However, the reality of the time course reactions demonstrate a gradual decrease in fluorescence measurements over time, particularly noticeable as reactions approach their plateaus. This finding indicates the presence of a second enzymatic rate in the observed reaction, and is indicative of DNAzyme degradation. Consequently, in light of this background rate of degradation, conventional percent cleavage calculations based on fluorescence generation in the cleaved and uncleaved bands do not accurately depict target-mediated DNAzyme cleavage. Furthermore, kinetic analyses must also be adjusted since the cleavage effect can no longer be accurately approximated by a pseudo-first order reaction model.

The effect of nuclease degradation might be mediated by proportional degradation of cleaved and uncleaved DNAzyme forms. However, gel analysis demonstrates unequal degradation of these DNAzyme forms, further complicating our ability to accurately estimate the ability of the candidate DNAzymes selected to detect their intended targets. This effect is illustrated in Figure 5, which depicts a comparison of the 12-1 DNAzyme performance in two MHS patient samples, with accompanying graph depiction of the fluorescence generated over time for the cleaved and uncleaved DNAzyme forms. Ideally, nuclease degradation could be discounted provided indiscriminate rates of degradation are demonstrated between the cleaved and uncleaved DNAzyme forms. This finding would manifest as equal and proportional concomitant decreases in fluorescence signal in the uncleaved bands with respective increases in signal in the corresponding cleaved bands.

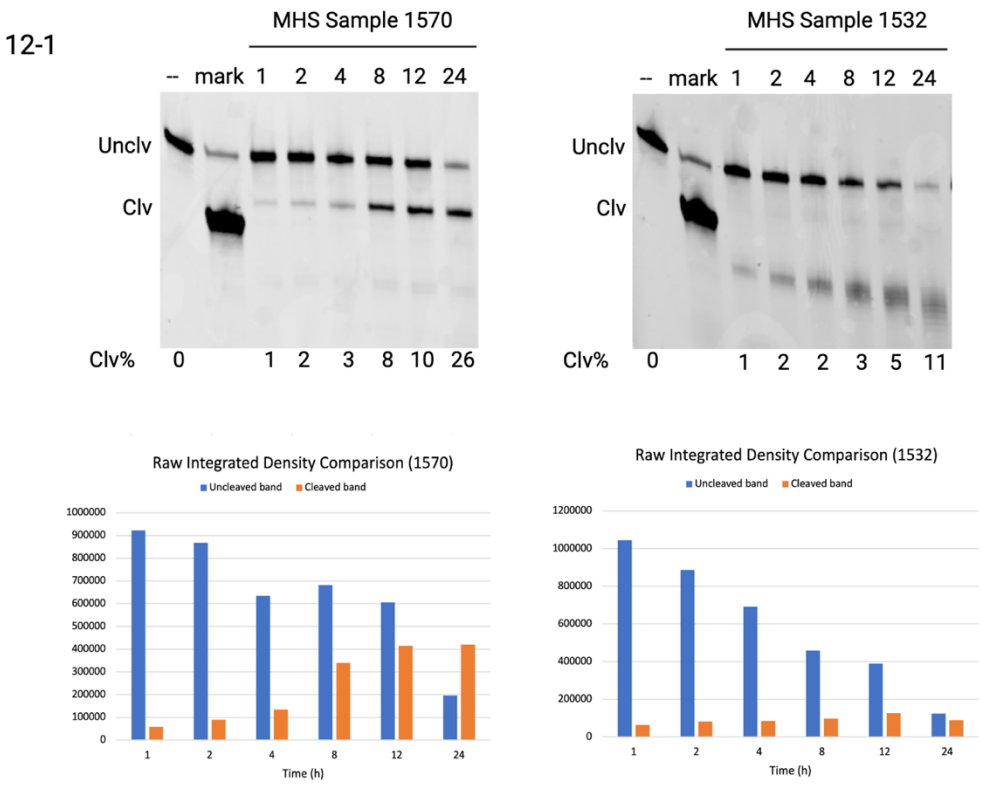


Figure 5. Nuclease degradation comparison of uncleaved and cleaved 12-1 DNAzyme forms. Gel scan and graph representation of fluorescent signal generation. Cleavage band raw intensity values precede the overamplification correction (division by 6) to further illustrate degradation discrepancies.

Figure 5 conclusively demonstrates unequal rates of degradation between the cleaved and uncleaved DNAzyme forms when comparing the raw integrated density values with the observed gel patterns. These raw integrated density values represent the measurement units of fluorescence emission and are generated by the fluorescence gel scanner. The discrepancies in nuclease degradation of the uncleaved and cleaved DNAzyme forms are particularly noticeable in MHS patient sample 1532. In this sample, the bar graph shows a precipitous decline in fluorescence emission of the uncleaved band; however, neither the gel image nor the bar graph demonstrate an expected increase in fluorescence emission in the cleavage band, suggesting the decline in fluorescence is attributable to non-specific degradation of the uncleaved band. Meanwhile, the gel

image and raw density measurements of the cleaved band show consistent values through 12 hours, with a small decline at 24 hours, indicating widely varying degradation rates between the two DNAzyme forms. This effect is more subtle in MHS patient sample 1570 but still detectable. A possible explanation for this degradation variation is offered by the steric hindrance effect of the fluorophore once released from being bound to the ribonucleotide cleavage site, conferring some protection against nuclease degradation.

Furthermore, both patient samples examined in Figure 5 shed light on the inability of the conventional cleavage percentage calculation to accurately approximate the percent cleavage at each time point. This effect is especially pronounced in later time points secondary to more extensive DNAzyme degradation over time and is apparent even to the naked eye in patient sample 1532. As previously noted, the cleavage band maintains equivalent raw intensity values over the course of 24 hours, with a small decline in fluorescence emission measurements at 24h. Yet, the conventional percent cleavage calculations suggest increasing cleavage values up to 11% by 24 hours. When compared to the intensity of the 10% cleavage band in sample 1570 (12h time point), the 24h cleavage band percentage in sample 1532 is evidently flawed, secondary to extensive degradation of the uncleaved DNAzyme band observed on the gel image, skewing the percent cleavage calculation towards a false positive value.

Similarly in patient sample 1570, the bar graph shows a gradual decline in fluorescence emission in the uncleaved band, in accordance with the observed effect in the correlated gel scan. While the cleaved band does demonstrate a converse increase in raw fluorescence density over time, it is not proportional to the decline demonstrated by the uncleaved band, indicating the presence of a

second reaction rate. This effect is particularly noticeable in the 24h time point of the gel image of sample 1570. At this time point, the percent cleavage calculation indicates the cleavage band holds 26% of the total DNAzyme signal for the time point. In the absence of degradation, the cleavage percentage value should correspond to the percent depletion of signal in the uncleaved band. Evidently, the uncleaved band appears to be significantly more depleted. In fact, when compared to the raw density values of the uncleaved marker and 1h uncleaved band, the 24h uncleaved band represents 15% and 21% of the comparison bands, respectively. This comparison leaves >50% of the DNAzyme fluorescence signal unaccounted for, and consequently is attributed to degradation by nuclease activity. Similarly, a comparison of the raw density values of the 12h and 24h time points show near identical values (415233 vs 420074, respectively), evidently not on par with the percentage calculations, once again demonstrating the skewed effect of the significantly more degraded uncleaved band on the cleavage percent calculation.

The results illustrated in Figure 5 therefore suggest that the conventional percent cleavage calculation is flawed when applied to complex biological matrices with more than one rate of degradation. Some consideration was given to methods of optimizing this calculation. A first option would see the addition of all fluorescence emission along the length of the well of each time point, in order to recreate the total DNAzyme signal and offer a more accurate ratio of the cleavage band fluorescence emission to the total signal present in the lane. However, given the extent of degradation observed across the selections over time, a vast majority of the degraded fragments will have run off the gel by the time adequate separation of the cleaved and uncleaved bands has occurred. Furthermore, it would be impossible to discern whether degraded fragments belong to the uncleaved DNAzyme form or the cleaved DNAzyme form, the latter of which requires

adjustment of the raw fluorescence density value to account for over-amplification of the fluorescent signal when the fluorophore is no longer in close proximity to the quencher. Therefore, this method would also suffer from considerable inaccuracy.

A second simple alteration to the percent cleavage calculation would see the use of a standard marker band not susceptible to degradation but identical in molar amount and concentration to the amount of DNAzyme in each time point, such that all percent calculations are normalized against this standard marker. While this method does not account for the effect of degradation on the cleavage band, it does eliminate the detrimental effect of the more profound degradation rate from the reaction.

In order to adequately subtract the nuclease effect from the calculations, a more comprehensive approach to data quantification needs to be considered. One such option includes the use of an extensive background subtraction. This method would involve the use of a non-cleaving all DNA control – an oligonucleotide sequence the same length and identity as the DNAzyme construct examined, but with the catalytic components removed such that its function is extracted. By incubating this sequence in a parallel serum time course reaction, a more accurate account of the percent degradation at each time point along with the degradation rate can be extracted and subsequently used to adjust the fluorescence emission values generated by the fluorescence scanner. This method would be especially helpful in correcting the observed rate constant of a DNAzyme to account for background degradation by plotting both the DNAzyme cleavage and the DNA control degradation over time and performing a background subtraction of the DNA

control rate from the DNAzyme cleavage observed rate, yielding a more accurate rate constant for the DNAzyme's activity in the sample tested.

Evidently, the most effective method of eliminating the nuclease degradation effect is to specifically neutralize the nucleases effect and prevent the degradation altogether. The ensuing discussion will enumerate suggested methods of inactivating/preventing nuclease degradation. Further studies are required to test these suggestions prior to incorporation into a standardized *in vitro* selection protocol. Beginning with nuclease inhibiting cocktails, these options include metal ion chelation options[5], as well as small molecule[6], protein[6], and antibody inhibitors.[7] Metal ion chelation-based inhibitor cocktails are evidently not viable for use with DNAzyme reactions since the nuclease inhibiting effect would extend to the DNAzyme's activity as well. Conversely, small molecule inhibitors are designed to bind to nucleases by mimicking substrate strands, and protein and antibody inhibitors are designed to reversibly bind and form complexes with nucleases, slowing their activity.[5,6] These may constitute viable options for control of nuclease degradation. However, prior to their use, they require screening against the DNAzymes being tested to ensure they do not affect DNAzyme activity by way of inappropriate interactions.

A second approach to controlling nuclease activity may come from the use of decoy DNA. In this approach, addition of tRNA or salmon sperm DNA can be titrated into the selection reactions to effectively saturate the nucleases in solution, staving off non-specific degradation of the DNAzyme library/candidates. Consequently, the DNAzymes in the reaction are spared to pursue meaningful interactions with targets of interest, improving the effectiveness of each selection reaction, especially counter selection.

A third approach could see depleting or eliminating nucleases from serum altogether. Given its drastic approach however, this method comes at the expense of altering the biological landscape of the serum sample employed. This method would involve flowing serum through a column with affixed DNA substrates for nucleases, such that nucleases bound to the column are eliminated from the serum and the flow through serum can be used to perform the subsequent selection reactions. This method could also be accomplished with incubation of DNA-affixed beads added to serum, followed by bead removal such that removal of the beads subsequently removes the nucleases from the remaining serum.

In conclusion, the detrimental effect of non-specific nuclease activity on the selections performed has been highlighted, along with multiple suggestions for ameliorating the damaging influence of nuclease activity on the efficacy of *in vitro* selections. Further studies are required to assess the efficiency of these suggestions, towards the development of a robust and standardized approach to *in vitro* selection in complex biological matrices such as human serum.

5.2.4 Optimization of *in vitro* selection protocol in human serum

A recurring theme throughout this thesis has been the need for development of an optimized approach to performing *in vitro* selection experiments in human serum. Given that biomarkers have risen to the forefront of human theranostic research, there has never been a greater need for the development of a new biomarker identification method, and *in vitro* selection is extremely well equipped to fill this role. By standardizing the approach to *in vitro selection* in whole human serum, the scientific community can benefit from the elimination of time-consuming challenges and pitfalls, making the pursuit of truth and knowledge more efficient. Consequently, this section is an accumulation of the lessons learned over the course of the two selections performed, in the hopes that they will serve as a launchpad for the development of a defined protocol for conducting *in vitro* selection experiments in complex biological samples.

Based on the results of the seroconversion method comparison discussed earlier in this chapter, it appears that the optimal serum derivation method is direct from whole blood, yielding unadulterated native serum. While nuclease degradation appears to be higher in this seroconversion method, the use of decoy DNA or nuclease depletion via bead or column approaches may prove fruitful in overcoming this challenge. However, consideration must be given to the limitation posed by access to patient samples. Consequently, if unable to access whole-blood derived serum, acquiring plasma followed by defibrination with thrombin and calcium addition appears to offer the most efficient approach to serum derivation.

With regards to sample age, the overall impression from the selections performed would suggest that collection age does not have a readily detectable negative impact on sample quality, provided

adequate storage practices of biological samples are followed. Furthermore, recent studies appear to contradict the previous belief of accumulated freeze-thaw cycles having a detrimental effect on the quality of biological samples.[8,9] However, particularly when investigating an unknown target, sound scientific practices would err on assuming a low threshold for target degradation, and therefore aim to conduct selection experiments on fresh collected samples, along with minimizing idle time at room temperature.

Turning our attention to the intrinsic selection protocol, several challenges encountered in the selections performed may be circumvented through fine-tuning the selection protocol when applied to human serum. For instance, given the difficulties experienced with counter selection, it may be worthwhile exploring the effect of introducing counter selection earlier in the protocol, or even beginning with counter selection, in order to readily eliminate cross-reactive DNAzymes from the very beginning of the experiment and limit the amplification of sequences undergoing non-specific serum interactions. While this suggestion comes at the expense of greatly diminishing the starting DNA library population, it may warrant exploration to assess its ability to improve the efficiency of counter selection. Increasing counter selection incubation time further or even increasing the ratio of counter selection rounds relative to each positive selection round may represent viable strategies for increasing the stringency of counter selection. Furthermore, increasing the serum concentration in counter selection reactions in order to maximize DNAzyme library interactions with common serum components may also prove to be an effective strategy to drive up the specificity of the DNAzyme library. Ultimately, the largest detrimental effect on the specificity of the DNAzyme candidates generated has been the effect of non-specific nuclease degradation. Effective neutralization of their detrimental effects with strategies discussed earlier

in the chapter would likely lead to a significant increase in the effectiveness of counter selection and the subsequent specificity of emerging DNAzyme candidates.

Similarly, strategies for improving sensitivity of DNAzyme candidates also warrant discussion, particularly given the low percent cleavage values observed in the sensitivity analysis of the new malignant hyperthermia samples. Promising strategies include shortening the positive selection incubation time and decreasing the serum concentration in positive selection reactions (in effect decreasing the target concentration), effectively prioritizing affinity as a selective pressure. Applying these pressures ensures the DNAzyme sequences with highest affinity for the target pursued are selected, increasing the sensitivity of the DNAzyme library

Furthermore, given that *in vitro* selection is a dynamic process, both positive and counter selection may benefit from addition of new patient samples over time rather than maintaining static positive and counter selection pools of serum. By introducing new samples to the protocol over the course of a selection, we can effectively introduce more heterogeneity reflective of the general population, thus increasing the generalizability of our findings while optimizing the sensitivity and specificity of emerging DNAzyme candidates.

5.3 Summary

In conclusion, we have identified multiple challenges faced when performing *in vitro* selection in serum. Despite initial experiments suggesting seroconversion method does not impact DNAzyme functionality, subsequent experiments and analyses have conclusively shown that serum preparation methods non-uniformly impact DNAzyme functionality. This determination underscores the importance of creating a standardized approach and protocol for *in vitro* selection in human serum, particularly regarding sample preparation, with whole-blood derivation of serum appearing to be the most promising and reliable method of seroconversion. We have additionally explored the detrimental effect of non-specific nuclease activity on the selections performed, and offered multiple strategies for addressing and anticipating nuclease activity prior to commencing *in vitro* selection protocols in complex biological samples. We have additionally provided protocol-specific optimization suggestions in the hopes of improving the efficacy of *in vitro* selection in complex media.

5.4 Experiments

5.4.1 ALS DNAzyme assay

The FQ30-ALS6 library construct, along with the FQ-30-R98 DNAzyme construct (derived from ALS6) was provided by the lab technician. The ALS6 library resembles the architecture of the PanC2 and MH1 libraries with regards to the presence of conserved primer regions of equal size flanking a 40-nucleotide random sequence domain. The library and DNAzyme were ordered from IDT and purified on 10%dPAGE prior to FQ30 ligation. Ligated constructs were subsequently purified on 10% dPAGE as well. The same formulation of 10x Selection Buffer was used (500mM HEPES pH 7.0, 1500mM NaCl, 150mM $MgCl_2$) for incubation reactions. The ligated library and DNAzyme were incubated with multiple blood fractions and serum variations for 4 hours. The cleavage was then analyzed by 10% dPAGE. The image of cleaved and uncleaved library and DNAzyme bands was obtained with the Amersham Typhoon Biomolecular Imager.

5.4.2 Plasma and serum preparation

ALS-positive patient blood samples previously collected by the ALS project collaborators in EDTA-coated vacutainers and processed to plasma were reprocessed to serum[10] using three methods: Calcium + Thrombin, Calcium only, and Calcium + Silica beads. In each method, 200µl of plasma was seroconverted using either 5µl of Calcium Chloride-Thrombin stock solution (0.025U/µl, 0.5M), 2.5µl 1M Calcium Chloride, or 2.5µl 1M Calcium Chloride and 20 silica beads. Clotting was achieved in each seroconversion method.

5.4.3 Calcium concentration and seroconversion experiment

Multiple concentrations of $CaCl_2$ stocks were prepared, as follows: 1M, 0.5M, 0.25M, 0.125M to facilitate reliable addition of calcium to small volumes of serum. Four MHS plasma samples used to create the Positive Selection Serum Pool were seroconverted using Thrombin as previously described, using a 200µl aliquot. In total, 8 serum samples were selected for this experiment, all of which were successfully seroconverted with both methods: addition of 1-3 doses of $CaCl_2$ +/- Silica beads, or with Thrombin addition. Additionally the samples were selected such that 4 samples came from the first shipment of samples, and 4 samples came from the second shipment of samples for further comparison. The samples were separated by conversion method. Serial concentrations of each serum sample were performed by addition of $CaCl_2$ doses to reach the desired concentrations of Calcium in serum (multiples of 12.5mM to 100mM). Aliquots of serum were taken after each Calcium dose administered for use in the experiment. Once the serum fractions were prepared, the ligated 12-1-FQ30 DNAzyme construct was incubated with each serum fraction for 24 hours, and cleavage time course reactions were performed as previously described in earlier chapters.

5.5 References

1. Dominguez K, Ward WS. A Novel Nuclease Activity that is Activated by Ca^{2+} Chelated to EGTA. *Syst Biol Reprod Med*. 2009;55(5-6):193-199. doi:10.3109/19396360903234052
2. Barra GB, Santa Rita TH, Vasques J de A, Chianca CF, Nery LFA, Costa SSS. EDTA-mediated inhibition of DNases protects circulating cell-free DNA from ex vivo degradation in blood samples. *Clin Biochem*. 2015;48(15):976-981. doi:10.1016/j.clinbiochem.2015.02.014

3. Balian A, Hernandez FJ. Nucleases as molecular targets for cancer diagnosis. *Biomark Res*. 2021;9(1):86. doi:10.1186/s40364-021-00342-4
4. Hirsh J, Warkentin TE, Shaughnessy SG, et al. Heparin and Low-Molecular-Weight Heparin Mechanisms of Action, Pharmacokinetics, Dosing, Monitoring, Efficacy, and Safety. *Chest*. 2001;119(1):64S-94S. doi:10.1378/chest.119.1_suppl.64S
5. Kolarevic A, Yancheva D, Kocic G, Smelcerovic A. Deoxyribonuclease inhibitors. *Eur J Med Chem*. 2014;88:101-111. doi:10.1016/j.ejmech.2014.07.040
6. Pasloske BL. Ribonuclease Inhibitors. In: *Nuclease Methods and Protocols*. Humana Press; :105-111. doi:10.1385/1-59259-233-3:105
7. Ohashi A, Murata A, Cho Y, et al. The expression and localization of RNase and RNase inhibitor in blood cells and vascular endothelial cells in homeostasis of the vascular system. *PLoS One*. 2017;12(3):e0174237. doi:10.1371/journal.pone.0174237
8. Lee JE, Kim SY, Shin SY. Effect of Repeated Freezing and Thawing on Biomarker Stability in Plasma and Serum Samples. *Osong Public Health Res Perspect*. 2015;6(6):357-362. doi:10.1016/j.phrp.2015.11.005
9. Gislefoss RE, Lauritzen M, Langseth H, Mørkrid L. Effect of multiple freeze-thaw cycles on selected biochemical serum components. *Clinical Chemistry and Laboratory Medicine (CCLM)*. 2017;55(7). doi:10.1515/cclm-2016-0892
10. Castro AR, Kikkert SE, Fears MB, Pope V. Defibrination of Blood Plasma for Use in Serological Tests for Syphilis. *Clinical and Vaccine Immunology*. 2002;9(6):1376-1378. doi:10.1128/CDLI.9.6.1376-1378.2002

Chapter 6. Conclusion and Outlook

This thesis research has explored the feasibility of performing *in vitro* selection in whole human serum towards the identification of DNAzyme sequences capable of disease detection. In addition to demonstrating the viability of this application, we have identified multiple DNAzyme candidates with promising diagnostic potential, and developed a bioinformatic analysis method for evaluating the progress of *in vitro* selection and identifying candidate sequences with high predicted sensitivity and specificity for a specific target.

To explore the feasibility of performing *in vitro* selection in whole human serum, two distinct pathologies were studied. In chapter 2, we focused on deriving DNAzymes specific to pancreatic cancer detection. Using patient samples purchased from the Ontario Tumor Board, we conducted the preliminary phase of the experiment, demonstrating that a naïve DNA library can be enriched in human serum towards developing functionality. The validation phase of the experiment yielded numerous DNAzyme candidates demonstrating consistent sensitivity towards pancreatic cancer, and two DNAzymes (14-4 and 14-15) capable of distinguishing pancreatic cancer from normal human serum in fresh collected whole-blood derived serum (native serum).

The next phase of this project would see a high-powered sensitivity and specificity analysis of the 14-4 and 14-15 DNAzymes to establish their sensitivity and specificity in native serum, and against other cancer types. Additionally, optimization of 14-4, 14-15, and the additional candidates highlighted in Chapter 2 through reselection, truncation, and structure manipulation may yield improved sensitivity and specificity for pancreatic cancer. Furthermore, sensitivity and specificity

analyses of each candidate DNAzyme with spiked naïve library may further improve their diagnostic performance, in addition to revealing valuable insight into the mechanism of action of distinct DNAzymes, in this case indicating whether any candidate sequences are activated by the presence of a secondary sequence to facilitate its folding or attachment to a specific target.

In chapter 3, we focused on a non-acute pathology to diversify the scope of *in vitro* selection in human serum and assess whether the feasibility demonstrated in Chapter 2 is reproducible. Using a similar protocol to the one employed in Chapter 2 and patient samples from the Malignant Hyperthermia Investigation Unit, we successfully completed the preliminary phase of the experiment. The results yielded an enriched DNA library, confirming the success of *in vitro* selection in human serum, and four candidate DNAzyme sequences which advanced to sensitivity and specificity analysis. Through the generosity of the MHIU, we were well equipped to perform a thorough sensitivity and specificity analysis, which concluded that all four candidate sequences hold clinically relevant sensitivity and specificity values across a range of target incubation times.

These preliminary validation results are very encouraging and warrant further clinical investigation. Furthermore, optimization of each candidate sequence using the strategies previously listed may further improve their ability to distinguish MH sensitive serum from MH negative serum. With further clinical validation, the DNAzymes identified in this selection can be included in the development of a biosensing platform for point-of-care diagnostic detection of MH susceptibility.

In chapter 4, we explored the breadth of bioinformatic analysis, substantiating its importance in guiding and evaluating the *in vitro* selection process. We additionally showcased the wealth of information which can be extracted from sequencing data, particularly as it pertains to identification of promising DNAzyme candidates and using sequencing data to characterize *in silico* kinetic performance. In doing so, we successfully identified an additional DNAzyme capable of acting as a general cancer detection probe. This discovery holds tremendous promise for the development of an inexpensive, minimally-invasive preliminary cancer screening tool.

Furthermore, we identified a two-step method for evaluating and informing next steps of the *in vitro* selection protocol. The first step involves identifying and tracking round-to-round enrichment of high-frequency sequences within DNA libraries recovered from each round of selection. The second step involves clustering of the sequences identified in the final selection round, and performing cluster-based analyses for identification of prominent candidates with specificity towards an intended target. Further collaborations with data and computer scientists would aid in furthering the depth of information we can access from sequencing data, using machine-learning algorithms to seek out further patterns among the sequencing data as it pertains to detection of functional motifs and predicted structure versus function analysis.

In Chapter 5, we conducted a critical assessment of the *in vitro* selection protocol implementation in human serum, highlighting challenges, discrepancies, and suggestions for improvement. This assessment uncovered the inconsistent effect different seroconversion methods have on DNAzyme functionality, the detrimental impact of non-specific nuclease degradation of DNAzyme probes on maintaining target specificity, and difficulty of conventional DNAzyme cleavage calculations to

accurately approximate target-mediated cleavage. Additionally, we demonstrated that library enrichment is enhanced with prolonged incubation time with positive selection serum, reflecting the expected likelihood of any target specific to a pathology of interest being present in very low concentrations. The challenges presented in this chapter were counteracted with strategies and suggestions for addressing and preventing their detrimental effects, in the interest of maximizing the efficiency of *in vitro* selection in human serum and any other complex biological matrices.

The fifth chapter aimed to inspire the future development of a standardized approach to *in vitro* selection in human serum. By creating a standardized protocol, the scientific community can benefit from the elimination of time-consuming challenges and pitfalls, making the pursuit of disease-specific functional nucleic acids and novel disease biomarkers more efficient. In doing so, we can ensure that diagnostic solutions to pathologies in desperate need of earlier recognition can come to fruition, in addition to being able to readily compare and reproduce results with confidence.

In conclusion, this thesis describes the ability to perform *in vitro* selection directly in whole human serum across widely differing pathologies. Our hope is for this work to inspire further application of this method in serum and other complex biological samples, towards the identification of novel disease biomarkers and functional nucleic acid-based detection probes. Furthermore, we hope the candidate sequences identified in this work may find themselves incorporated into diagnostic platforms, offering minimally invasive, reliable, point-of-care diagnostic tools for early diagnosis and prevention of disease.