



Precision Nucleic Acid Diagnostics using Novel Oligonucleotide Reagents and Amplification-Free Assays

Slott, Sofie

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Slott, S. (2023). *Precision Nucleic Acid Diagnostics using Novel Oligonucleotide Reagents and Amplification-Free Assays*. DTU Chemistry.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Precision Nucleic Acid Diagnostics using Novel Oligonucleotide Reagents and Amplification-Free Assays

Sofie Slott

PhD Thesis

January 2023

Principal Supervisor: Associate Professor Kira Astakhova

Co-supervisor: Professor Mads H. Clausen

Department of Chemistry
Technical University of Denmark
Kemitorvet, Building 206
2800 Kgs. Lyngby
Denmark

Preface

The work presented in this thesis is the result of my PhD studies from November 2019 to January 2023 carried out at the Department of Chemistry, Technical University of Denmark. The PhD studies were conducted under the supervision of Associate Professor Kira Astakhova as main supervisor and Mads H. Clausen as co-supervisor. The project was financed by the Department of Chemistry, Technical University of Denmark through the PhD Excellence program.

A six-months external stay was carried out in the Ji Research Group, Department of Medicine, Division of Oncology at Stanford University, from October 2021 to March 2022 under the supervision of Professor, M.D., Hanlee P. Ji. The external stay was partially funded by Fulbright Denmark.

The thesis is divided into 6 chapters. Chapter 1 describes the research background of the thesis with the use of nucleic acid biomarkers in precision diagnostics, detection methods and issues. Chapter 2 summarizes current challenges and defines the motivation and outline of the projects in the thesis. Chapter 3 and 4 describe the work done with the FLEET platform. Chapter 4 is a submitted article. Chapter 5 relates to production of multiple probes at once. Finally, Chapter 6 covers the work done during my external stay at Stanford University related to oligonucleotide-selective sequencing as targeted sequencing method.

PhD Isabela Ferreira, MSc Asmund Kjellegaard Ottesen, and BSc Marie-Louise Wibom, Anna Kyung Jakobsen and Cecilie Krüger-Jensen all contributed to the work described in Chapter 3 and 4 concerning the FLEET assay. Isabela did the mathematical calculations and predictions using the PB mesoscopic model. Asmund, Marie-Louise and Anna helped with fabrication of chips and Cecilie with optimizations of LOD. It has been noted throughout the thesis when work was performed by others, and contributors have been credited by name.



Sofie Slott

Kongens Lyngby, January 31, 2023

Abstract

Precision in cancer diagnostics is of high interest to improve survival rates, costs and treatment guidance. Nucleic acid biomarkers are promising tools for use in precision cancer diagnostics. However, detection of nucleic acid biomarkers remains a challenge. This thesis will address some of the issues related to detection of nucleic acid biomarkers for early cancer diagnostic.

In the thesis, three strategies to optimize current amplification-free FLEET platform, established in Astakhova group, have been investigated. We applied calculations from the Peyrard-Bishop model to make insertions of LNA in high-affinity capture sequences targeting oncogenes, and predicted their melting temperatures (T_m). Selected probes were synthesized, and had their T_m measured, resulting in an accuracy of 1°C. This indicated that the PB model can be applied for future design of ultra-specific probes. Secondly, we implemented the FLEET assay in microfluidic PMMA chips using TC-tagged capture probes immobilized with UV light. We fabricated microfluidic PMMA chips and estimated LOD to be 6.4pM – 32pM using a spectrofluorometer. The assay was tested with BRAF-probes, designed with the PB model. This showed a detectable signal, but needs to be optimized.

We describe a quantitative detection method for mutated microRNA in human plasma samples. Specific oligonucleotides designed from a Peyrard-Bishop model allowed accurate prediction of target:probe recognition affinity and specificity. The assay allowed identification of single-nucleotide polymorphism mismatch profiles in clinically relevant microRNA-128-2-3p, showing terminal mutations that correlate positively with inflammatory colitis and colorectal cancer.

Related to production of probes, we describe a new approach to make pools of microRNA targeting breast cancer cells. The microRNA pools were synthesized at once on the same solid-support using ‘Tandem Oligonucleotide Synthesis’ strategy. We make up to four consecutive microRNA using 2’/3’OAc nucleotide phosphoramidites. The developed phosphoramidites combined give a cleavable moiety that separates the microRNAs and are cleaved using standard cleavage conditions. Furthermore, we investigate making branched pools (microRNA dendrimers) versus linear pools as a strategy to further improve the product yields. Our approach provides microRNA pools in high yields, which is of relevance to the growing demand on synthetic RNA oligomers for nucleic acid research and technology.

Targeted sequencing has great importance in finding new variants of nucleic acid biomarkers. In this thesis, we optimize *Oligonucleotide-Selective Sequencing*, established in the Ji Research lab. We investigate the effect of primer spacing using synthesized branched oligonucleotide as primers, with the aim of improving yield and sensitivity. The branching primers were found to improve sensitivity by 2-fold compared to control primers which is promising.

Resume

Præcision i cancerdiagnostik har stor interesse for klinikere og patienter, da det kan føre til forbedret overlevelsesrater, reducere omkostning og bedre individuel behandling. Nukleinsyre-biomarkører fra blodet har stort potentiale til brug i præcis cancerdiagnostik. Detekteringen af nukleinsyre-biomarkører er dog fortsat udfordrende. Denne afhandling vil behandle nogle af de problemstillinger relateret til detektering af nukleinsyre-biomarkører i tidlig cancerdiagnostik.

I denne afhandling, er tre strategier til optimering af nuværende FLEET-platform, etableret af Astakhova gruppen, blevet undersøgt. Ved brug af Peyrard-Bishop modellen beregnede vi de optimale placeringer af LNA i design af sekvenser med høj affinitet for onkogener, og beregnede derefter deres T_m . Udvalgte sekvenser blev syntetiseret og fik derefter målt deres T_m , hvilket resulterede i høj nøjagtighed på mindre end 1°C . Dette viste at PB-modellen kan anvendes til fremtidig design af ultra-specifikke prober. Dernæst implementerede vi FLEET metoden i mikrofluid PMMA-chips ved brug af TC-mærkede prober, der blev immobiliseret på PMMA overflade med UV-lys. Vi fremstillede mikrofluid PMMA-chips og estimerede LOD til at være mellem 6.4pM og 32pM ved brug af spectrofluorometer. Vi testede metoden med PB-designet BRAF-prober, som viste et detekterbart signal. Medvidere, fik vi udviklet en ny metode til kvantitativt at detektere mutationer i microRNA fra humane plasma prøver. Specifikke oligonukleotider blev designet ud fra PB-modellen, som ydermere lavede nøjagtige bestemmelse af target:probe genkendelsesaffinitet og specificitet. Assayet tillod identifikation af enkelt-nukleotid mismatch-profiler i klinisk relevante microRNA-128-2-3p. Dette viste terminale mutationer, der korrelerer positivt med inflammatorisk colitis og tyktarmskræft.

Relateret til produktionen af prober, beskriver vi en ny metode til at lave puljer af miRNA rettet mod brystkræft. Puljer af miRNA blev syntetiseret på én gang, på den samme support ved brug af *Tandem Oligonukleotid Syntese*. Vi lavede op til fire sammenhængende miRNA ved hjælp af 2'/3'OAc nukleotidphosphoramiditter, der kombineret danner en kløvbar funktionalitet, der kan spaltes ved brug af standard kløvningbetingelser. Desuden undersøgte vi forgrenede puljer af miRNA (dendrimerer) versus lineær puljer som en strategi til at forbedre udbyttet. Vores metode gav miRNA-puljer med høje udbytter, hvilket er af relevans for den voksende efterspørgsel på syntetiske RNA-oligomerer til nukleinsyreforskning og teknologi.

Måltrettet sekventering har stor betydning for opdagelser af nye varianter i nukleinsyre-biomarkører. I denne afhandling optimere vi *Oligonukleotid Selektiv Sekventering* metode, der er udviklet i Ji Forskningslaboratorie. Vi undersøger effekten af afstand mellem primere ved hjælp af syntetiserede forgrenede oligonukleotider som primerer for at forbedre udbytte og følsomhed. De forgrenede primere viste sig at forbedre følsomheden for metoden 2-fold sammenlignet med kontrolprimerne, hvilket er et lovende resultat.

Acknowledgements

First, I would like to thank my supervisor Associate Professor Kira Astakhova for giving me the opportunity to do a PhD study in your group, and for the great guidance throughout the project. I admire your ability to think big, being creative and always taking the positive side. It has been a great pleasure working with you.

I would like to thank Professor Hanlee P. Ji, for hosting my external stay at Stanford University and to all the people in the Ji group for making me feel very welcome and including me in the group activities. It was an absolute pleasure to be in your lab for 6 months, and I really enjoyed my time at Stanford both personally and professionally. I also want to thank GiWon Shin for supervising me in the lab and for being patient with me.

Furthermore, I would like to thank all the group members in KAS group, past and present, for creating a good working place and a good atmosphere: Adrian, Nadia, Vinzenz, Emilie, Pâmella, Izabela, Weiguang, Jonas, Anders, Sangita, Maria and Jesper. A special thanks to Adrian Hernández Bustos for helping me synthesizing oligonucleotide for the project, but also for being an amazing friend and colleague always ready to help. Thank you for your dance moves, attitude and music that always create a dance-party in the lab. I really appreciate it. I would also like to thank the KQ group for being good office mates and for always sharing snacks, cake and providing big screens in the lab for important moments. A special thanks to Amalie Nørskov for proofreading my thesis, but also for being a great friend and mental support through my PhD study.

Thank you to the students I have supervised. A special thanks to Asmund, Marie-Louise, Anna and Cecilie for a good collaboration and for the chip related work presented in this thesis.

I would like to extend my thanks to collaborating groups: Professor Gerald Weber and his group members for a great collaboration throughout the years. Especially thank you to Izabela Ferreira for the computer work provided, and for teaching me about mesoscopic models. Thank you to Martin Dufvas group at DTU Health Tech for helping with chip fabrication, trouble shooting and guidance.

Thank you to Mads H. Clausen for always being available and asking good question in Monday Morning Meeting. Thanks to the current and former technical staff at DTU chemistry for keeping anything running and always being helpful when assistance is needed. Thank you to David Frej for helping with special orders and biology skills. Thank you to Charlie Johansen for taking care of the MALDI-TOF and UPLC-MS. Thank you to all remaining technical staff, former staff, administrative staff, the building center, the NMR center, the reception and the IT-support. Thank

you to the Department of Chemistry for awarding me the Academic Excellence Scholarship to fund my PhD studies.

A special thanks to Jens Ø. Duus and Charlotte Held Gotfredsen for always having the door open for a great conversation, advice when needed or guidance during the years. I really appreciate it a lot, so thank you.

I would like to thank my frisbee-friends in KFK, for making a great community with lots of fun, great atmosphere, many tournaments and endless dance parties. It has given me so much energy in my daily life, and therefore really grateful for. Thank you to the Stanford women ultimate teams, Superfly and Firefly, for including me in the Stanford ultimate family and taken me to tournaments around the US. It made my external stay unforgettable.

Lastly, I would like to thank my family and friends for always supporting me in life and have my back no matter what. I am really grateful for having you all in my life, and I could not have done this without your support. Thank you to my sweet friends for always being up for spontaneous ideas and post-practice kebab. Thank you to my parents for all your love and support in life. Thanks to my sister Cille for being the best sister in the world, and to Emil for always being helpful, caring and providing quality jokes. Finally, thank you Nikolaj for your love, support and patience with me, especially in the last few months. I am truly grateful for having you to share the experience with, your always positive and rational point of view, and your ability to encourage me on a rainy day. It means the world to me, so thank you.

Thank you, I am truly thankful for having you all in my life.

List of Abbreviations

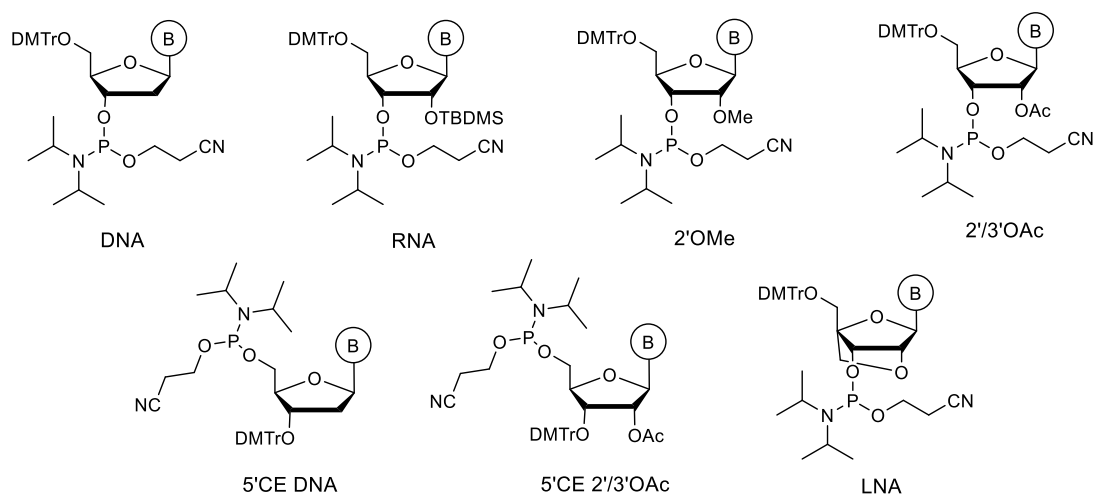
Ac	Acetyl
AC	AccuClear
AO	Acridine Orange
APEX	Attachment-based primer extension
ASO	Antisense oligonucleotide
bp	Base pair
Bt	Biotin
Cb	Crowding branched oligonucleotides
CD	Crohn's Disease
CE	2-Cyanoethyl
CPG	Control pore glass
CRISPR	Clustered regularly interspaced short palindromic repeats
CS	Cleavable site
CTD	Calf thymus DNA (“booster”)
ctDNA	Circulating tumor DNA
DCE	1,2-Dichlorethane
DCI	4,5-Dicyanoimidazole
DCM	Dichloromethane
ddPCR	Digital droplet PCR
DMSO	Dimethyl sulfoxide
DMTr	Dimethoxy trityl
DNA	Deoxynucleic acid
EG	EvaGreen

eq.	Equivalents
ESR	Erythrocyte sedimentation rate
EtOAc	Ethyl acetate
EtOH	Ethanol
FLEET	Synonym for ' <i>moving very fast</i> '
Hept.	Heptane
HPLC	High Performance Liquid Chromatography
HRMS	High resolution mass spectrometry
IEDDA	Inverse electron demand Diels-Alder cycloaddition
i5	Index 5 (barcode for multiplexing)
LNA	Locked nucleic acid
MALDI-TOF	Matrix-assisted laser desorption-ionization Time-of-flight
Me	Methyl
MeCN	Acetonitrile
min	Minutes
miRNA or miR	microRNA
MMT	Monomethoxytrityl
MQ	Milli-Q water
MS	Mass Spectroscopy
mut	Mutation
NGS	Next-Generation Sequencing
NHS	N-Hydroxysuccinimide
NMR	Nuclear Magnetic Resonance
NN	Nearest Neighbor model
NSCLC	Non-small cell lung cancer

nt	Nucleotide
OS-Seq	Oligonucleotide-Selective Sequencing
PB	Peyrard-Bishop mesoscopic model
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PMMA	Poly(methyl methacrylate)
PNA	Peptide nucleic acid
PoC	Point-of-care
PS	Phosphorthioate
PSA	Pressure Sensitive Adhesives
P5 primer	Primer on Illumina flow cell
P7 primer	Primer on Illumina flow cell
QF	QuantiFluor
qPCR	Quantitative PCR (also referred to as “Real Time-qPCR”)
R_f	Retention factor
RNA	Ribonucleic acid
RP	Reverse Phase
rt.	Room temperature
RT-qPCR	Reverse Transcription qPCR
Sb	Spacing branched oligonucleotide
SCLC	Small cell lung cancer
scr	Scramble
siRNA	Short interfering RNA
SNP	Single nucleotide polymorphism
SP1 primer	Primer for read 1

SP2 primer	Primer for read 2
TBDMS	<i>tert</i> -Butyldimethylsilyl
TBE	Tris/Borate/EDTA buffer
TCA	Trichloroacetic acid
TCO	<i>trans</i> -Cyclooctene
TEAA	Triethylammonium acetate buffer
T_m	Melting temperature
TO	Thiazole Orange
TONS	Tandem oligonucleotide synthesis
Tz	Tetrazine
UC	Ulcerative Colitis
UHPLC	Ultra High Performance Liquid Chromatography
USER	Uracil-Specific Excision Reagent Enzyme
WGS	Whole-Genome Sequencing
wt	Wild type

Phosphoramidite Monomers in Thesis



Contents

PREFACE.....	I
ABSTRACT	III
RESUME	V
ACKNOWLEDGEMENTS.....	VII
LIST OF ABBREVIATIONS	IX
1 RESEARCH BACKGROUND	1
1.1 Precision in cancer diagnostics	1
1.2 Nucleic acids	2
1.3 Nucleic acids in cancer diagnostic	3
1.3.1 Circulating tumor DNA.....	5
1.3.2 Circulating microRNA.....	5
1.4 Current detection methods of nucleic acid biomarkers.....	7
1.4.1 RT-qPCR	7
1.4.2 High throughput Sequencing.....	8
1.4.3 Microarray-based assay	9
1.4.4 Amplification-free methods	10
1.5 Probe design for hybridization assays.....	12
1.6 Oligonucleotide synthesis	14
1.6.1 Other synthesis strategies for making oligonucleotides	16
2 STUDY HYPOTHESIS, METHODOLOGY AND RATIONALE	17
2.1 Current Issues and State of the Art	17
2.2 Aim of Project 1: A diagnostic platform for early detection of cancer	18
2.3 Aim of Project 2: Tandem RNA Synthesis	20
2.4 Aim of Project 3: PCR-Free targeted single primer capture for ultra-deep sequencing.....	21
3 A DIAGNOSTIC PLATFORM FOR EARLY DETECTION OF CANCER.....	23

3.1	Hybridization detection assay: FLEET	23
3.2	Peyrard-Bishop mesoscopic model for probe design	24
3.3	Project outline and goals	26
3.4	Results.....	27
3.4.1	Use of Mesoscopic Model to Design Ultra-Specific Probes.....	27
3.4.2	FLEET using Microfluidic Chips	30
3.5	Conclusion and Future work	35
3.6	Experimentals.....	36
3.6.1	Oligonucleotides.....	36
3.6.2	FLEET assay on chip	40
3.6.3	T _m Measurements.....	42
3.6.4	Oncogene Information	42
4	MUTATIONS IN MICRORNA-128-2-3P IDENTIFIED WITH AMPLIFICATION-FREE HYBRIDIZATION ASSAY	43
4.1	Introduction	43
4.2	Experimentals.....	45
4.2.1	Model	45
4.2.2	Oligonucleotides.....	45
4.2.3	UV Melting Studies	45
4.2.4	Fluorophore Study in Chip.....	45
4.2.5	Cohort Information.....	46
4.2.6	miRNA Analysis.....	46
4.2.7	Statistical analysis.....	48
4.3	Results.....	48
4.3.1	“Booster” dye optimization	48
4.3.2	SNP probe design.....	49
4.3.3	Tandem hybridization bead-based assay – Detecting SNPs in human plasma miR-128-2-3p.....	51
4.3.4	Correlation with other clinical parameters	54
4.4	Discussion.....	54
4.5	Conclusions.....	56
	Author Information.....	57
	Author Contribution.....	57
	Conflicts of interest.....	57
	Acknowledgements	57
4.6	Supporting Information	58

4.6.1	Mesosopic Modelling.....	58
4.6.2	Oligonucleotides.....	62
4.6.3	UV Melting Studies.....	63
4.6.4	Dye study in chip.....	65
4.6.5	Patient Sample Information.....	69
4.6.6	miRNA Analysis.....	71
4.6.7	Statistical analyses.....	76
5	MICRORNA POOLS SYNTHESIZED USING TANDEM SOLID-PHASE OLIGONUCLEOTIDE SYNTHESIS.....	77
5.1	Tandem Oligonucleotide Synthesis.....	77
5.2	Results and Discussion.....	79
5.2.1	Study Design.....	79
5.2.2	Synthesis of Monomers.....	82
5.2.3	Synthesis of miRNA Pools.....	83
5.2.4	Pool Characterization.....	84
5.3	Conclusion.....	88
5.4	Experimentals.....	89
5.4.1	General Experimental Methods.....	89
5.4.2	Synthesis of Monomers.....	89
5.4.3	Synthesis of miRNA pools.....	94
5.5	Supporting Information.....	95
5.5.1	NMR-spectra of S1-S8.....	95
5.5.2	Oligonucleotide Design, Synthesis and Characterization.....	106
6	PCR-FREE TARGETED SINGLE PRIMER CAPTURE FOR ULTRA-DEEP SEQUENCING.....	111
6.1	Oligonucleotide-Selective Sequencing (OS-Seq).....	111
6.1.1	Principals of OS-Seq.....	111
6.1.2	Second Generation of OS-Seq on Agarose Beads.....	113
6.1.3	Sequencing parameters.....	115
6.2	Project outline: Optimization of OS-SeqV2.....	116
6.2.1	Design of branched oligonucleotides.....	117
6.3	Results and Discussion.....	120
6.3.1	Synthesis of Branched Oligonucleotides.....	120
6.3.2	Optimization of OSSeqV2 (2.1).....	122
6.3.3	Summary of OS-SeqV2 optimizations.....	126
6.3.4	Future work.....	126

6.4	Experimental	127
6.4.1	General Chemical Experimental Methods	127
6.4.2	General Methods for Molecular Assays	127
6.4.3	Oligonucleotide Synthesis	127
6.4.4	Gel Characterization of Branched Oligonucleotides.....	128
6.4.5	NHS coupling	129
6.4.6	OSSEQ-APEX Method.....	130
7	CONCLUSION	135
8	FUTURE PERSPECTIVE	137
	REFERENCES.....	139
	APPENDIX.....	1
A1	List of Publications and Conferences.....	1
	Publications.....	1
	Conferences	2
A2	Co-Author Statements	3
A3	Complete Mesoscopic Parameterization of Single LNA Modifications in DNA Applied to Oncogene Probe Design.....	10

1 Research Background

1.1 Precision in cancer diagnostics

Cancer is the second leading cause of death worldwide, accounting for nearly 10 million deaths in 2020¹. The most common types are breast cancer, lung cancer, colon and rectum cancer and prostate cancer. Lung cancer is the most lethal with 20% survival rate over 5-year life span¹. The 5-year survival rate for early-stage lung cancer (I) is 70-90%, however most patients (approx. 75%) have advanced disease at time of diagnosis (III/IV)². Despite development in oncologist management of late stage cancer, survival rates remain poor.² Early diagnosis of cancer is therefore important for improving survival rates². However, diagnosis of cancer currently suffer from low sensitivity, because many tumors cannot be found at early stage and delay treatment before it is too late.³

Personalized medicine or *precision medicine* is a growing demand within cancer diagnostics^{2,4,5}. Most medical treatments are designed for the “average patient” as a one-size fits all. Precision medicine is an approach that takes into account differences in people’s genes. The same disease can behave differently from one patient to another due to specific mutations in the genome^{6,7}. This can result in good response for some patients and strong response from others on the same treatment⁷⁻⁹. Precision medicine aims to target the right treatment for the right patient. Therefore, precision in cancer diagnostic is a key to localize specific mutations and guide treatment^{2,4,5}.

Recently it was discovered that non-invasive ‘liquid biopsies’ contain genomic biomarkers enable discrimination between various cancer patients from healthy individuals such as lung, gastric and colon cancer¹⁰⁻¹³ making them tools for earlier diagnostics³. Genomic material from liquid biopsies can reveal type of mutation that can guide the treatment, provide information about current tumor burden as well as identify specific mutations that arise during treatment^{3,8,14}.

Precision in cancer diagnostics is therefore of high interest and important in order to improve survival rates, costs and treatment guidance^{3,8,14}. This thesis will be focusing on nucleic acid biomarkers in precision cancer diagnostics, current issues and development of detection methods.

1.2 Nucleic acids

Nucleic acids are the fundamental components in all living cells. Natural nucleic acids are deoxyribonucleic acid (DNA) that store genes in the nucleus while ribonucleic acids (RNA) are transcripts of the coding DNA that have a regulatory role or translated into proteins in the ribosomes. There are different types of RNA, all with different functions in the cell. Messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA) are necessary for protein translation, whereas non-coding RNA such as microRNA (miRNA) and interfering RNA (iRNA) are regulators in the cell¹⁵. Nucleic acids consist of deoxy- or ribose sugar rings bound together with phosphate, where every sugar ring has a base attached to it (**Figure 1.1**). For DNA the four bases are Cytosine (C), Guanine (G), Adenosine (A) and Thymine (T) where RNA has the same except for Uridine (U) instead of thymine. They bind together antiparallel by Watson-Crick base pairing to its complementary base (A:T and C:G) to form double stranded helix structure¹⁶. The intermolecular forces in nucleic acids are hydrogen bonding and van der Waals forces obtained from π stacking between the base pairs¹⁷.

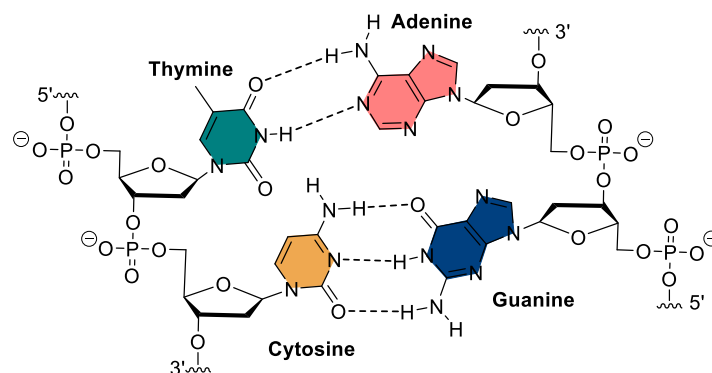


Figure 1.1: Watson-Crick base pairing between A:T and C:G.

A recent paper by Nikitin expands the traditional Watson-Crick model with so-called “communication” systems¹⁸. In general, communication stands for not-fully matched (“mismatched”) double stranded nucleic acids. Nikitin shows that these not fully matched duplexes can have biological role, being as short as only 7 nucleotides (nt) in length. Communication explains not fully complementary duplex formation for natural DNA and RNA and opens a path to new exiting discoveries in this direction.

According to traditional Watson-Crick model, mismatches are abnormal in DNA genome. When that happens, the incorrect base in the genome is exchanged with the right one with repair-enzymes for DNA¹⁹. However, for RNA, it is much more tolerated with mismatches when forming duplexes. They can bind to a target without being completely complementary²⁰. This is because many types of RNA has a natural 3D structure, where parts of the RNA are hybridized

together, and plays a role in the function of the RNA²⁰. Types of secondary structures in RNA are helices, hairpin loops, bulge loop, internal loop and junctions (**Figure 1.2**)²¹. Long RNA molecules can have many secondary structures forming tertiary structures, such as ribosomal RNA, while small RNA molecules like miRNA (20-22nt) usually appear as hairpin or loops and allow for mismatch when bound to target^{20,21}.

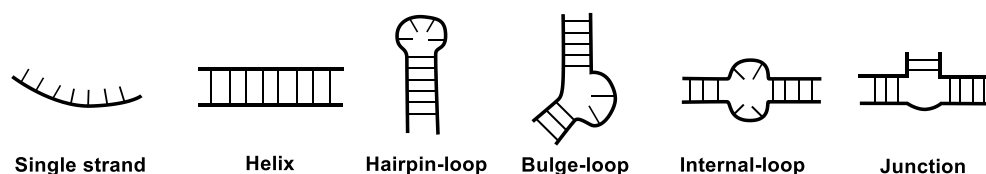


Figure 1.1: Secondary structures of RNA. Helix, hairpin loop, bulge loop, internal loop and junction¹⁵.

1.3 Nucleic acids in cancer diagnostic

Cancer is a genetic disease and is caused by alterations in the genes that are controlling cell growth and cell division²². During cell division, genetic errors can happen and mutations in the DNA can occur. Some genetic mutations are not harmful and the cell is able to repair. However, when cell proliferation happens frequently; naturally, it is more prone to get mutations that can have influence for the cellular processes. Genetic mutations that causes cancer can be inherited or arise from inflammatory environmental exposure such as tobacco, UV radiation or chemical toxins²². Lung cancer is the most lethal accounting for 18% of all cancer deaths¹. It is known that tobacco smoking is the most important risk factor for developing lung cancer, responsible for 80% of lung cancers²³. There are two types of lung cancer: Small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). SCLC account for approximately 15-25% of all detected lung cancers whereas NSCLC accounts for 75-85%². NSCLC is more widespread, however, SCLC is more aggressive, fast growing and metastasize more easily².

Single-nucleotide polymorphism (SNPs) are variations of a specific nucleobase on the genome level²⁴. In cancer disease, SNPs are usually found in oncogenes, meaning that the mutation is causing a change in favor of the cancer. SNPs can be “helpful tools” in cancer diagnostic because they are sequence - and disease specific, which makes it possible to compare a cancer genome with healthy genome only looking at SNP regions. Most common oncogenes for NSCLC are *KRAS*, *EGFR* and *BRAF* mutations found in 30%, 20% and 5% respectively in all NSCLC cases²⁵. *EGFR* encodes for the epidermal growth factor receptor, important for cell replication. Mutation within this receptor like L858R, is associated with lung cancer and cause overexpression of the gene resulting abnormal cell proliferation⁶. *KRAS* encodes for the K-Ras protein involved in signaling regarding cell proliferation. K-Ras respond to EGFR activation. Common SNPs in *KRAS*

are G12D and G13D which cause cell proliferation and growth²⁵. *BRAF* gene encodes for B-Raf proteins importing for directing the cell growth²⁵. *BRAF* V600E mutation is present in 1-2% of all NSCLC and cause uncontrolled cell division and growth²⁵.

Early detection of cancer is essential for improving survival rates²⁶. This is because early symptoms can be difficult to recognize. Liquid biopsies such as blood, urine, feces and saliva, have become an important role in early cancer diagnostic because cancer-specific genetics and abnormalities in certain biomarkers can be revealed^{3,8,13,27}. Furthermore, liquid biopsies are non-invasive, which is huge advantage for fast and easy monitoring patients. Different biomarkers exists in a blood sample such as proteins, metabolites, circulating tumor cells (CTC), circulating tumor DNA (ctDNA), cell free DNA (cfDNA), circulating microRNA and extracellular vesicles (**Figure 1.3**)²⁸. The nucleic acid biomarkers are interesting because they contain cancer-specific information that can guide the optimal treatment for the individual patient. Therefore, circulating tumor DNA and circulating microRNA have the potential to improve personalized medicine⁵.

Current challenges with early cancer detection can be divided into five: 1) Need for better understanding of biology in early disease. 2) Determine risk factors for disease. 3) Finding good biomarkers for early disease. 4) Development of accurate technologies for detection. 5) Evaluating early detection approaches appropriately²⁶. Furthermore, there is a general need for accurate early detection technologies that address issues of cost, access and scaling, and integration with public health infrastructure²⁶. Point-of-care testing (PoC) is diagnostic testing of patients in the clinic such diagnosis can be established rapidly. PoC tools have emerged during COVID-19 pandemic and may also help advance the implantation of early cancer detection²⁶. The development of new approaches to integrate sample preparation of nucleic acid detection, such as ctDNA and miRNA, in a faster, cheaper and user-friendly format still remain desirable and challenging²⁹.

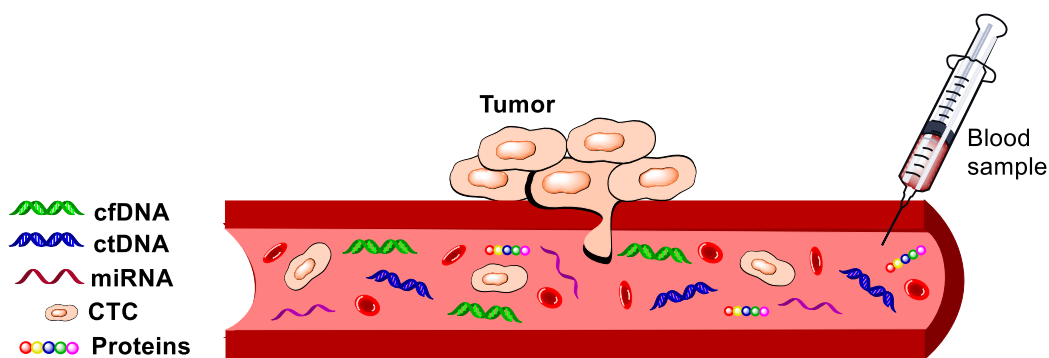


Figure 1.3: Liquid biopsy with ctDNA and miRNA as biomarkers. ctDNA and miRNA are released from tumor site into blood stream. cfDNA are released from normal cells. Other biomarkers circulating are circulating tumor cells (CTC), proteins, extracellular vesicles and metabolites.

1.3.1 Circulating tumor DNA

Circulating tumor DNA (ctDNA) are small DNA fragments released from the tumor site into the blood stream³⁰. This is happening due to abnormal growth in the tumor, where cells are dying and replaced by new ones. In this process, the apoptotic or necrotic cell releases DNA into the blood and become cell free. ctDNA contains SNPs that are oncogene-specific such as mutations in *BRAF*, *EGFR* or *KRAS* and can therefore be used as biomarkers. However, when normal cells are dying, the same process is happening and DNA is released into the blood stream and is known as cell-free DNA (cfDNA). The length of ctDNA can vary, having a median fragment length is 134-144bp, which is around 20-30bp shorter than cfDNA³⁰. Moreover, ctDNA accounts for only 0.1-10% of the total circulating cell-free DNA, whose normal plasma levels range from 10-100 ng/ml³¹ but increase in ctDNA level with progression of disease³². Therefore, method of detection has to be very sensitive and specific for catching such small amounts. Current detection methods of ctDNA are targeting methods such as digital droplet PCR and Next-Generation Sequencing methods which are described in the next section²⁸.

1.3.2 Circulating microRNA

microRNA (miRNA) are small non-coding single stranded RNA molecules in length of 20-22nt and function as regulators of gene expression in cells. They bind to complementary messenger RNA (mRNA) in the cell thereby preventing the translation of mRNA to proteins through degradation or repression. miRNA are transcribed from DNA in the nucleus to primary miRNA (pri-miRNA) which are processed into precursor miRNA (pre-miRNA) and transferred to cytoplasm where they are matured to miRNA³³. Mature miRNA can then bind to mRNA through the RNA-induced silencing complex (RISC) and induce silencing or degradation (**Figure 1.4A**). MicroRNA are the cells natural tool to up- and down regulate genes and they can have different function depending on the local environment of a specific miRNA. A single miRNA can regulate hundreds of different mRNAs, while many different miRNA molecules can target the same mRNA. miRNA also exists as circulation miRNA in the blood where they have shown to be very stable and functioning as signaling to other cells³³.

Over the past decades it has become clear that miRNA expression is dysregulated in human cancer cells.³⁴ Genome investigations has revealed that many miRNA genes are located in cancer associated genomic regions³⁴. This suggest that abnormal miRNA expression in cancer cells could arise from amplification or deletion of specific genomic regions encompassing miRNA genes³⁴. This result in miRNA typically being up- or down regulated in the favor of the cancer. miRNA targeting oncogenic mRNA are downregulated (tumor-suppressor miRNA) whereas miRNA targeting tumor-repressor mRNA is up-regulated (oncogenic miRNA) (**Figure 1.4B**). An example

of oncogenic miRNA is miR-21, that has been shown to be overexpressed in many cancer types such as breast, lung and colon cancer^{35–37}.

Dysregulation of miRNA play an important role in cancer development³³. It has recently been showed that miRNA have a key function in cancer cell metastatic spread³⁸. Therefore, miRNA are both interesting biomarkers for diagnosis and prognosis of cancer. The abnormal expression can distinguish between healthy patients and cancer patients. Due to their role in cancer development, miRNA has recently become therapeutic targets for cancer treatment^{39,40}. By making synthetic miRNA that either mimic the downregulated tumor-repressor miRNA or inhibit oncogenic miRNAs (anti-miRNA) miRNA can be used as therapeutic targets (**Figure 1.4 C+D**)^{39,41,42}. Furthermore, it has been verified by next-generation sequencing, SNPs in miRNA correlate with their biogenesis and are linked to inflammatory and cancerous diseases²⁴. This happens when SNPs introduce structural changes into miRNA impairing their function and thereby modulate gene expression level, causing disease susceptibility or pathogenesis^{43–45}. In the light of growing interest for SNP profiling in miRNAs, accurate and sensitive detection methods are on high demand²⁴.

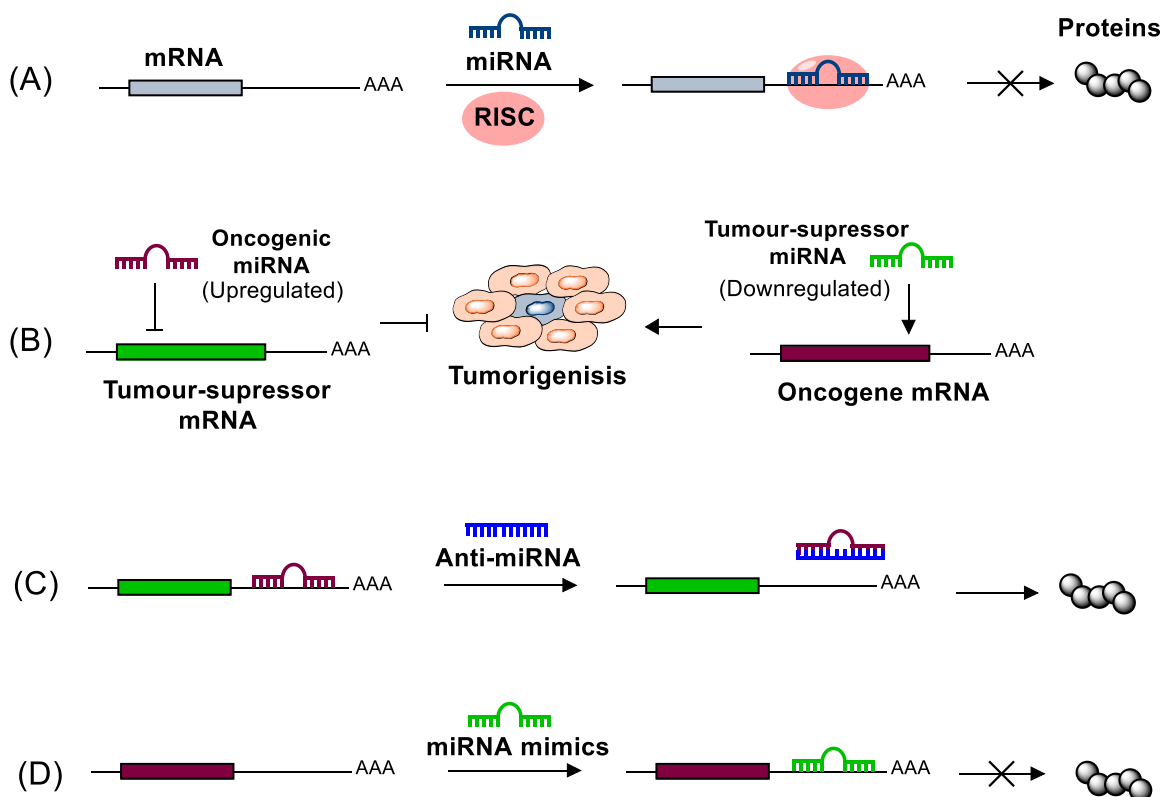


Figure 1.4: Function of miRNA in normal cells, cancer cells and as therapeutics. (A) miRNA function to degrade or inhibit translation of mRNA to proteins. (B) miRNA act either as oncogenic miRNA (upregulated) which target tumor-suppressor mRNA or as tumor-suppressor miRNA (downregulated) that act on oncogenic mRNA. (C) Mechanism of action for anti-miRNA therapeutics. (D) Mechanism of action for miRNA mimics therapeutics.³³

1.4 Current detection methods of nucleic acid biomarkers

Clinical sensitivity refers to the ability of a test to identify those patients *with* disease, whereas clinical selectivity refers to identification of those *without* disease. Challenges for detection of nucleic acid biomarkers lays within analytical sensitivity and selectivity. Analytical sensitivity refers to limit of detection (LOD), whereas analytical selectivity or specificity is how good the method is to discriminate between targets. Due to the low abundance of ctDNA in complex media and miRNAs short sequences and ability to fold, both sensitivity and selectivity are important for in detection⁴⁶. Current methods of detecting are RT-qPCR, Next generation sequencing (NGS) and microarray-based hybridization assay circulating microRNA^{3,46}, and RT-qPCR with ddPCR and targeted NGS methods for ctDNA⁴⁷. The methods are briefly described below with and are summarized in **Table 1.1** at the in the end ⁴⁶.

1.4.1 RT-qPCR

Reverse transcriptase quantitative PCR (RT-qPCR) is the golden standard for detection of microRNA and used as validation for many new developed methods (**Figure 1.5A**)^{48,49}. For microRNA it requires, as the name indicates, an initial revers transcription to coding DNA (cDNA), followed by quantitative PCR amplification (qPCR). The latter is done by adding a sequence specific probe with fluorescent tags (TaqMan) which hybridize to the coding strand, and every time the polymerase is making a copy, the probe is quenched, and a fluorophore molecule is released. The fluorescent signal is monitored each cycle and the quantity can be determined. qPCR can also be made with SYBR Green, which is an intercalating dye that interact with DNA, and then the fluorescence is determined after each cycle and quantified. RT-qPCR is cost-effective, well establish method with high accuracy, however it does require conversion of microRNA to cDNA and amplification steps⁴⁶. Moreover, primer design can be difficult when targets are small, such as miRNA, because mismatch can occur resulting in false positive read-out. New miRNA cannot be detected using this method, therefore mainly used for validation. In a Meta-study, Yi et al. collect data of all miRNA detected using RT-qPCR or microarray to determine miRNA biomarkers in NSCLC⁵⁰. qPCR can also be combined to optimize NGS, such as Saelee et al. that quantify high molecular weight DNA in cfDNA using qPCR to adjust ctDNA input amount for optimal NGS assay performance⁵¹.

1.4.1.1 Digital droplet PCR

Digital droplet PCR (ddPCR) is a PCR method, where a sample is portioned out in emulsion of oil and water micro-droplets⁵². In each droplet, qPCR is carried out with TaqMan and based on the fluorescent signal each droplet is assigned positive or negative. Using ddPCR it is possible to

remove non-target signals and improve selectivity of the method⁴⁶. ddPCR is time consuming and cost of reagents are high, and special equipment is needed⁴⁶. Never the less, ddPCR is still wildly applied for detection of ctDNA due to high sensitivity^{47,53,54}. Recently, ddPCR was applied to detect EGFR mutations in ctDNA from NSCLC patients, and ctDNA mutations in ovarian cancer follow-up patients^{53,54}.

1.4.2 High throughput Sequencing

Next-Generation Sequencing (NGS) is a massively parallel sequencing technology that offers ultra-high throughput of genomic material. It is for both circulating miRNA and ctDNA where new variants are found with high accuracy (99.99%)⁵⁵. Many sequences platform applied sequencing by synthesis (SBS) methods. This imply genomic DNA being fragmented and prepared with adaptors containing primers (P5, P7, SP1,SP2) and unique index barcodes (i5)⁵⁶. This allow for multiplexing, complement binding to flow cell primers, bridge amplification and sequencing. In **Figure 1.5B**, an example of single-index sequencing using Paired-End Illumina Flow Cell are shown⁵⁶. NGS methods are never the less expensive, time consuming and generate an exceptional amount of data, therefore not suitable for large scale diagnostic applications nor clinics⁴⁶. Moreover, RNA needs to be converted to cDNA and amplification is needed for library prep of the sequencing where PCR errors can happen⁴⁶. Recently, Yoshinami et al. applied NGS and barcoding to sequence gene panels in breast tumors and ctDNA in plasma from stage I or II patients⁵⁷. This showed mutations in ctDNA can be detected with high sensitivity in early breast cancer patients⁵⁷. New NGS methods for detection of miRNA biomarkers in liquid biopsies are also developed due to the demand of precision diagnostic from non-invasive biopsies⁵⁸.

1.4.2.1 Targeted NGS methods for diagnostic

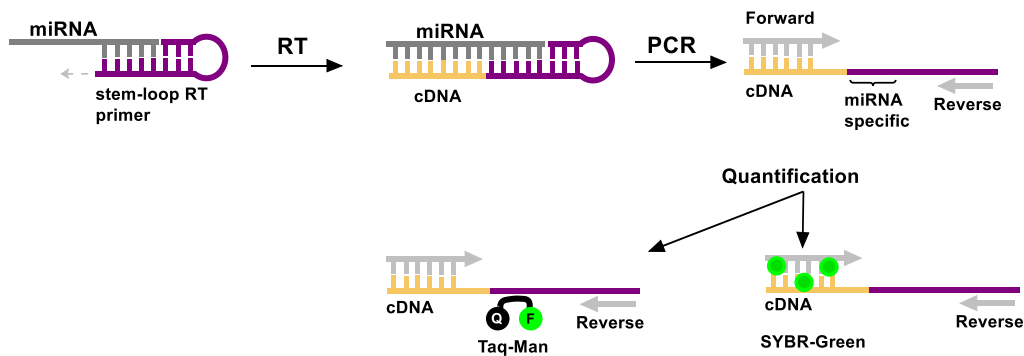
Targeted sequencing can be used instead of NGS as a cheaper and faster alternative. In targeted sequencing or deep sequencing, only targets of interest are examined and therefore less time consuming and more suitable for diagnostic purposes. This imply looking for targeted variations such as SNPs, a single gene or a panel of genes. Different types of deep sequencing methods have been developed such as cancer personalized profiling by deep sequencing (CAPP-seq) and Safe-Seq^{59,60}. In CAPP-Seq, special probes are made to target known mutations in ctDNA which are sequenced. Kato et al. use CAPP-Seq to analyze ctDNA to identify resistance mechanisms in osimertinib-treated patients with EGFR T790M-positive NSCLC⁶¹. In Safe-Seq, a unique identifier (UID) is assigned to each DNA template, which is amplified and sequenced⁶⁰. This method was designed to improve sensitivity of NGS and has been applied to ctDNA of patients with metastatic colorectal and gastrointestinal stromal tumors for tracking therapy response with

specificity of 98.9%^{62,63}. In this thesis, oligonucleotide-selective sequencing (OS-Seq) is applied as targeted sequencing method and discussed in Chapter 6.

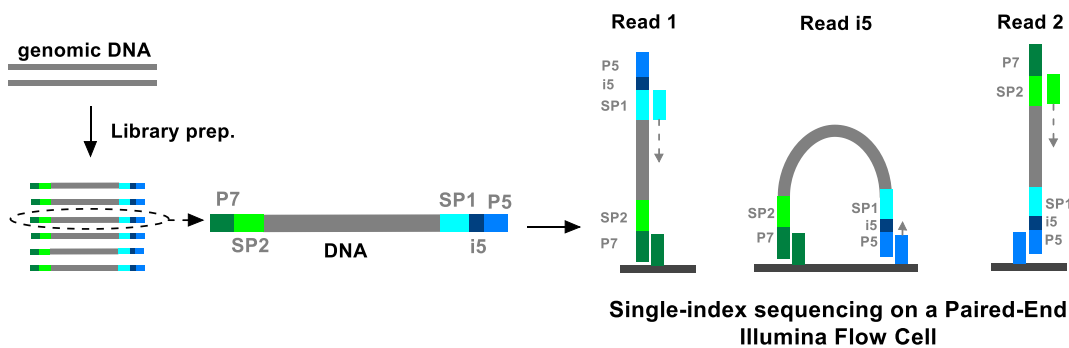
1.4.3 Microarray-based assay

Microarrays are hybridization-based biochips, where surface immobilized DNA probes bind target miRNA to obtain a fluorescent signal (**Figure 1.5C**). First microRNA is reverse transcribed into cDNA and labeled with sample specific fluorophore. Next, the fluorescent probes are added to the chip where complement probes bind and light up. Microarrays are rapid, high-throughput and cheap method to detect microRNA in tumor and healthy samples. The problem of this method is issues with mismatch possibilities resulting in false positive (low selectivity) and new miRNA variants cannot be detected with microarrays. Yet, microarrays are still widely applied to detect miRNA as biomarkers, such as in Zhang et al. identifying 28 different miRNAs associated with lung cancer⁶⁴ and Liang et al identified hsa-miR-21b, hsa-miR-29b and hsa-miR-155, miR-10b, hsa-miR-125b and hsa-miR-145 as biomarkers for breast cancer⁶⁵.

(A) RT-qPCR



(B) Next-Generation Sequencing



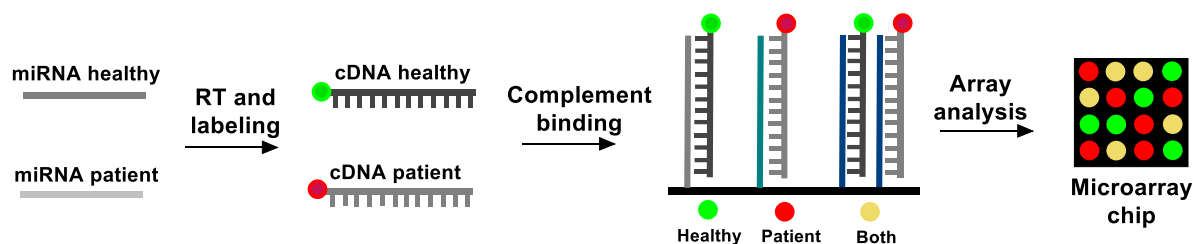
(C) Microarray detection

Figure 1.5: Current detection methods for miRNA and ctDNA from liquid biopsies. (A) RT-qPCR: miRNA is converted to cDNA (RT) and quantified by qPCR, using TaqMan or SYBR Green⁴⁹. (B) NGS: DNA library is fragmentation and ligated to adaptors with sequencing primers and index. Library is sequenced using SBS on e.g. a Paired-End flow cells from Illumina⁵⁶. In this SBS, adapters bind to P7 primers on flow cell. Then read 1 primer (SP1) is added to generate first read, followed by binding of P5 primer initiating read i5 and final addition of SP2 primer will generate read 2⁵⁶. (C) Microarray: miRNA from healthy of patient samples are extracted individually, reverse transcribed into cDNA and labeled with sample specific fluorophores. Fluorescent cDNA are added to microarray for complement binding. Red dots are miRNA from patient samples, green from healthy and yellow if both are bound.

In summary, RT-qPCR is the golden standard for miRNA and ctDNA detection, however probe design of small miRNA targets remain to be a problem and results in mismatch and false positive⁴⁶. Accuracy of NGS methods are very high, but is rather expensive and not suitable for point of care testing. Microarrays has potential for being PoC but lack specificity and sensitivity. Furthermore, all three methods require PCR amplification prior to detection which can affect content of nucleic acid in sample and give errors⁴⁶.

Methods	Specificity	Sensitivity	Advantage	Disadvantage
RT-qPCR	Middle	High	Automatic quantification	Medium-throughput and cannot identify new miRNAs
Microarray	Low	Low to middle	Low-cost, high-throughput, fast	Cannot identify new miRNAs
NGS	Very high	Middle	High-throughput and WGS	Expensive and long-run

Table 1.1: Current detection methods applied for detection of miRNA and ctDNA.⁴⁶

1.4.4 Amplification-free methods

Amplification-free methods are promising as they avoid sample amplification and therefore more suitable for point-of-care detection. Several methods have been designed to overcome specificity

in miRNA detection, but only few of them are SNP selective^{66–68}. The use of PCR amplification is to ensure high sample concentration, sufficient to be above the limit of target detection. In amplification-free methods, it is necessary to amplify the signal instead of the sample, to reach similar sensitivity level. This can be done using different approaches. Qiu et al. described the use of hybridization chain reaction (HCR) to fluorescently identify miRNA (let-7a) on SNP level using two different silver nanocluster hairpins reaching LOD of 1.7 nM⁶⁶. Similar, Jia et al. recently applied a CRISPR/Cas12a RNA directed nuclease system, in combination with HCR, for reverse transcription-free and amplification-free miRNA detection⁶⁹. Here the HCR circuit was used to bind Cas12a/crRNA complex, enable cleavage of FQ-labeled nucleic acid reporter giving a fluorescent signal. The method was not sensitive to SNPs, but reach limit of detection of 100 fM.⁶⁹

Another way to signaling binding of target miRNA is to use label capture probes. Caputo et al. designed a miRNA detection platform using microgels and molecular beacon stem loops, that when conjugated together, could catch miRNA in samples⁶⁷. In this method, each stem-loop is equipped with a fluorophore (Atto647N) that only is activated when target miRNA is bound. The binding of several miRNA at once amplifies the signal, and the biosensor showed SNP selectivity and low sensitivity (nanomolar to picomolar order, LOD 10fM, detection time 1h). Instead of amplifying the signal, Roy et al. used gold nanoparticles conjugated probes in microarray like approach to target miRNA. Target specific hairpin capture probes were immobilized on carboxyl-PEG functionalized glass slides to capture target microRNA. When microRNA was bound, Au-NP-linker was added and the array imaged using differential interference contrast (DIC) microscopy, which offers single-molecule resolution reaching limit of detection of 10fM⁶⁸.

Despite successful detection of miRNA and discrimination, these methods are still either expensive in materials and reagents, require enzymes and special detection or equipment are needed. Therefore, it is important to investigate in less expensive methods for use in point-of-care detection.

1.4.4.1 Microfluidic chips and materials

Microarrays are made on glass slides where capture probes are immobilized covalently. The SiOH groups on glass surface are modified with either an electrophilic or nucleophilic functional group, such the covalent bond can be established⁷⁰. Several materials have been investigated as cheaper options for glass slides, such as agarose films, polydimethylsiloxane (PDMS) and poly(methyl methacrylate) (PMMA)^{71–73}. Furthermore, different immobilization methods using these surfaces has been investigated⁷⁴. It has been shown that T₁₀C₁₀-tail DNA probes (TC-tags) can be immobilized inexpensively to both PMMA, agarose and glass surfaces using UV light successfully^{71,72,75}. The mechanism is not described, but it is believed that thymine react with the carbonyl in PMMA via a photochemical reaction such as [2+2] cycloaddition to make a stable conjugation. The idea of using plastic surface such as PMMA in combination of UV-light as

immobilization source is very appealing because it has the potential to lower the cost of current microarray detection dramatically. Therefore, it is a potent material for potential use in PoC devices.

1.5 Probe design for hybridization assays

When designing probes for capturing nucleic acid biomarkers like miRNA and ctDNA, sequences have to be highly selective and well designed to be successful. They need to bind to the right target, but also discriminate between sequences on a single base pair level, to avoid false positive. Making a rational design of these probes is a challenging task, because when two complementary strands are in close contact they will spontaneously bind to each other through Watson-Crick base-pairing, even if they do not match completely. This tendency is more prevalent for single stranded RNA which allow more mismatches than for DNA¹⁵. When targeting SNPs in oncogenes or miRNA the capturing probe has to cover the SNP, placing the SNP in the middle or 5'-end but can be flexible (**Figure 1.6**)⁷⁶.

In probe design, it is essential to look at melting temperature (T_m) of the target:probe duplex. The T_m will be higher for a fully match sequences compared to one containing a mismatch. Therefore, this can be used as a tool when designing the probes. Several parameters affect the melting temperature of a duplex. Length of the hybrid increase T_m of the duplex. A unique probe match is set to be minimum 16nt for being specific for its target⁷⁷. GC content does also affect T_m due to the extra hydrogen bond compare to AT. Salts in terms of buffers (Na^+ , K^+ and Mg^{2+}) are stabilizing the helix by reducing the repulsive Coulomb interactions of the phosphate groups and neutralizing the negative charged backbone⁷⁸. Increase in salt concentration will increase T_m , however, a higher concentration of monovalent ions (Na^+ and K^+) compared to divalent ions (Mg^{2+}) is needed in order to maintain the same T_m of a duplex⁷⁹.

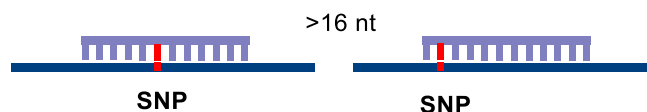


Figure 1.6: Probe design for SNP in oncogenes. Probe is designed either placing the SNP centrally or in the 5'-end but can be flexible, with minimum of 16nt overlay to be unique ^{76,77}.

Another way to increase the T_m between probe:target is by modifying the probe sequence with nucleic acid analogues to increase binding affinity for target (**Figure 1.7**). Locked nucleic acid (LNA) was first time described independently by Jesper Wengel and Takeshi Imanishi in 1997 and is one of the most applied nucleic acid modification in industry to increase binding affinity

to target sequences^{80,81}. LNA is locked in a C3'-endo/N conformation reducing its flexibility, resembling an RNA helix. LNA has stability towards nuclease enzymes and show increased specificity and affinity towards target probe⁸². LNA enrichment in probes can increase T_m up to 10°C per modification, and enhance mismatch discrimination⁸³. However, LNA cannot be inserted anywhere because it will affect 3D folding and does not have stabilizing effect anywhere in the probe⁷⁶. Other bridge nucleic acid analogues that have proven strong affinity for RNA target are tricyclo-DNA (tcDNA)⁸⁴, constrained ethyl (cET)⁸⁵ and ethylene nucleic acid (ENA)⁸⁶ which all have been used as ASO therapy candidates⁸⁷. Peptide nucleic acid (PNA) is also a class of nucleic acid analogues that has increased affinity to DNA^{88,89}. The backbone of PNA consist of amide bonds replacing the ribose-phosphate backbone, making the PNA sequence un-charged. This result in PNA binding to DNA without electrostatic repulsion making a stable helix with high affinity. PNA has been applied as hybridization probes in many detection methods to capture miRNA, such as in Oxford Nanopore Sequencing or signal amplification platforms using AuNP^{90,91}. This thesis will focus on LNA enrichment for probe design.

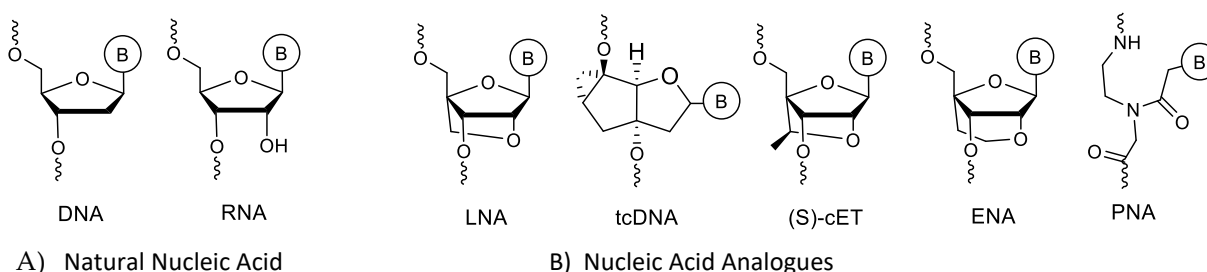


Figure 1.7: Nucleotide modifications used for stabilizing DNA:RNA duplex. Deoxyribonucleotide (DNA); ribonucleotide (RNA); locked nucleic acid (LNA); tricyclo DNA (tcDNA); constrained ethyl (cET) and peptide nucleic acid (PNA).

Despite the extensively use of LNA the last two decades, the origin of its enhanced base-pairing stability is still not sufficiently understood. It has been suggested that increase in hydrogen bonding and stacking interactions may both play a role in the enhanced stability effect of LNA, however conflicting explanations has been given⁹²⁻⁹⁵ which calls for more studies.

The most common method to design probes with high T_m is by using in silico methods to predict melting temperatures. Most online tools for predicting T_m for nucleic acid sequences are using Nearest Neighbor (NN) models^{93,95-97}. The NN model predicts DNA thermodynamics using energy values (ΔG° , ΔH° , ΔS°) for the different base pair motifs. These values have been derived from melting experiments in monovalent and divalent salt and applied to predict melting temperatures of oligonucleotides within a few degrees⁹⁸. However, these NN model provide little

insight into the intramolecular interactions such as hydrogen bonding and stacking interactions and unable to indicate the physical source of the increased LNA stability⁹⁹.

1.6 Oligonucleotide synthesis

Short synthetic nucleic acids called oligonucleotides are mainly synthesized using automated oligonucleotide solid-phase synthesis with phosphoramidite chemistry. Marvin Caruthers developed this method in the early 1980s in the wake of the invention of the solid-phase chemical synthesis by Bruce Merrifield in the 1960s^{100–102}. Since then, oligonucleotide solid-phase synthesis has revolutionized the field of biotechnology, therapeutics and diagnostics and is applied today to make gene therapy such as siRNA and ASO in industry.

Oligonucleotides are synthesized on solid-support in a 3'-end to 5'-end direction using phosphoramidite monomer building blocks that are attached one by one through a cycle of 4 reactions (**Figure 1.8**). These building blocks are protected nucleosides with a phosphoramidite group (2-cyanoethyl-*N,N*-diisopropylphosphoramidite) on the 3' position, dimethoxytrityl group (DMTr) at 5' position, and a protecting group on the 2' in case of RNA nucleosides. The solid-support (Control Pore Glass or Styrene) is equipped with a cleavable linker upon addition of a basic reagent (ammonia or methylamine) release the oligonucleotide.

Each cycle in the automated process consists of four reaction steps: 1) Deprotection; 2) Activation and Coupling; 3) Capping and 4) Oxidation (**Figure 1.8**). Every time a new nucleotide has to be added to the sequence, the cycle is repeated. The synthesis is initiated by deprotecting the hydroxyl group on the solid-support using trichloroacetic acid (TCA, 3% in DCM) liberating the orange DMTr group. Next, the nucleoside phosphoramidite is activated using an imidazole or tetrazole like compound as catalyst such the hydroxyl group can attack and make the coupling between nucleoside and linker. Standard phosphoramidite couplings are usually very efficient with coupling efficiency close to 99.9%. However, it is not possible to reach 100% in each cycle, therefore unreacted 5'-OH groups are capped with acetic anhydride (CapA) and *N*-methylimidazole (CapB). This will avoid deletion mutation sequences generated from potential reaction with next added nucleoside phosphoramidite. In the final step, the generated phosphite-triester (III) is oxidized to phosphate (V) using iodine in presence of pyridine and water. This is done to convert the acid labile phosphite-triester into a stable phosphate group prior to the TCA deprotection. After oxidation, the cycle is repeated until the desired sequence is achieved. To cleave the final sequence from solid-support, basic conditions in presences of heat are applied which also removes protecting groups from nucleobases and the 2'-cyanoethyl protecting group from phosphate.

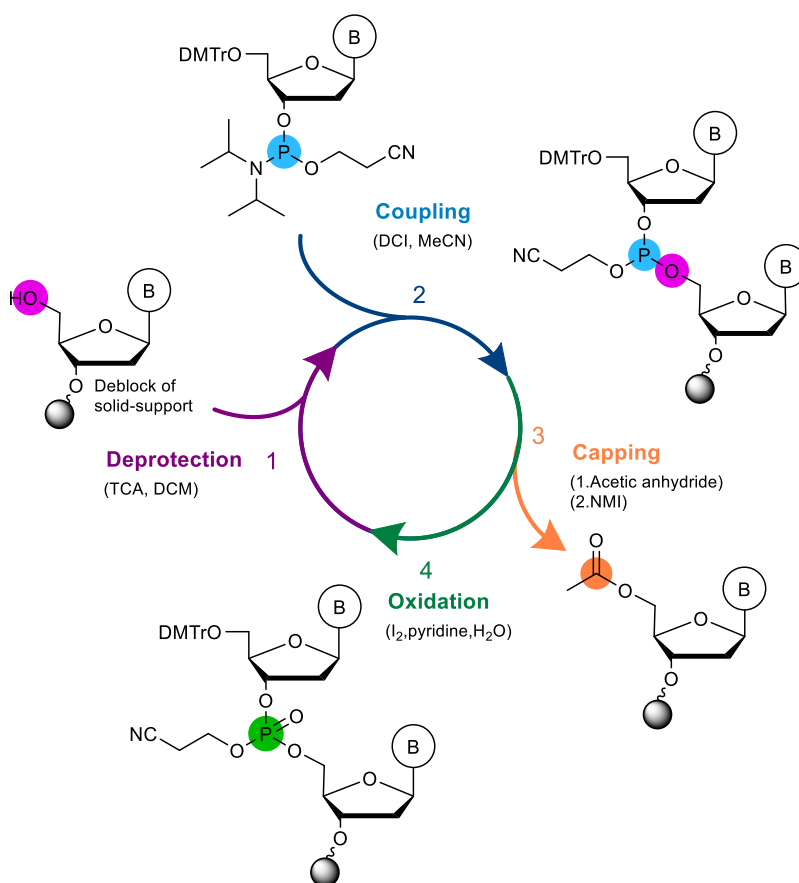


Figure 1.8: The cycle in automated oligonucleotide solid-phase synthesis. The four steps in the cycle: (1) Deprotection, (2) activation and coupling, (3) capping and (4) oxidation.

DNA-, RNA- and modified nucleoside building blocks such as 2'OMe, LNA and phosphorothioate (PS) linkages can be incorporated in the automated process to generate modified oligonucleotides. Furthermore, modifiers in 3'-end or 5'-end can be integrated prior to bioconjugation and 'click' reactions to functionalize the oligonucleotide. Most oligonucleotides synthesized using automated solid-phase synthesis are 20-30nt in length, such as siRNA, ASO, microRNA and primers. Guide RNA (gRNA) for CRISPR purposes are longer (up to 100nt) and can also be synthesized using this method. However, for a 50mer, 100mer and 200mer will result in 61%, 36% and 13% yield respectively even with 99% coupling efficiency in each step. Depurination can be an issue for longer sequences due to the use of acid in the deprotection step¹⁰³. Furthermore, RNA phosphoramidites do not couple as efficient and fast as the DNA phosphoramidites, therefore more challenging to make long. Another limitation to the solid-phase method is the challenge of making high throughput parallel synthesis, where multiple different sequences made at the same time¹⁰⁴. Even with years of optimizations, this challenge has

still not been solved. Never the less, this method is by far the most applied for making oligonucleotides on big scale for therapeutic purposes.

1.6.1 Other synthesis strategies for making oligonucleotides

Microarray oligonucleotide synthesis was developed to synthesize multiple oligonucleotides at the same time. It was developed in the early 1990s by Fodor et al. and later improved by Nimblegen and Gao et al. and is based on phosphoramidite chemistry¹⁰⁵⁻¹⁰⁸. The reaction takes place in a small droplet on array, where thousands of individual probes are made simultaneously. These phosphoramidite monomers uses photocleavable deprotection instead of acid to avoid depurination^{108,109}. This has been very useful in the biotech and research world where many different primers (DNA size ~20nt) are used for enzymatic reactions. Furthermore, pools for different screening purposes can be made, using this strategy. The advantage of this platform is the high throughput of oligonucleotides that can be made simultaneously (fmol scale). However, the method is limited to research and technology since the big number of probes makes it almost impossible to do quality control on the individual oligonucleotide and cannot be purified¹¹⁰.

2 Study Hypothesis, Methodology and Rationale

2.1 Current issues and State of the art

In summary, there is a need for early detection of cancer related biomarkers to improve survival rates². Nucleic acid biomarkers like miRNA and ctDNA can be found in blood, which allows for non-invasive liquid biopsy for monitoring patients and potentially, guide of treatment²⁸. Challenges for current detection of miRNA lays within the short sequence and tendency to fold in 3D structures resulting in mismatch and false positives signals. ctDNA has longer sequence, but occur in very low abundance among cfDNA. Therefore, high selectivity and sensitivity are both important for detection of nucleic acid biomarkers.

Current detection technologies include PCR methods, NGS methods and microarrays. These methods have different advantages and limitations. Some of the limitations are requirement of PCR amplification, reverse transcription of RNA samples before analyzing, cost, time and advanced equipment⁴⁶. Furthermore, health care and patients would benefit from cost-effective solutions that also could be a take-self test²⁶. Amplification-free hybridization assays have the potential to be a point-of-care technology. However, signal-amplification usually require many enzymes, fluorescent-labeled probes, advanced reagents and equipment^{66–69}.

Probe design for ultra-specific target recognition is essential in hybridization assays to avoid false positive signals. Important factors in probe design are melting temperature (T_m), GC-content, salt and nucleic acid modification such as LNA, to increase binding affinity to target. Most often, Nearest Neighbor models are applied to predict melting temperature of capture:target duplex^{93,95,96}. However, this model predicts T_m from estimated thermodynamic values based on existing T_m datasets. They do not provide info regarding hydrogen bonding and stacking interactions that are suggested to be important factors of the improved stabilizing effect of nucleic acid analogues⁹⁹.

Oligonucleotide solid-phase synthesis using the phosphoramidite method is the preferred way of making oligonucleotides in industry and applied for synthesis of gene therapy. Yet, there are

no applied method to make several oligonucleotides on the same support, even though most RNA therapy formulations need at least 2 sequences. Microarrays are used to parallel synthesize thousands of oligonucleotide at the same time in fmol scale in research and biotechnology. However, despite high-throughput, it is close to impossible to purify and make quality control of single oligonucleotide using this method.

The research conducted in this thesis consist of overall three projects concerning the matters stated above and are described below with hypotheses and project outlines.

2.2 Aim of Project 1: A diagnostic platform for early detection of cancer

Some of the issues in precision nucleic acid diagnostics lay within design of ultra-specific target recognition probes, point-of-care options for early detection of NA biomarkers and detection of SNPs in miRNA^{26,28}. The Astakhova Group has previously developed an amplification-free hybridization based assay called *FLEET*, to capture miRNA and ctDNA in sera samples^{111,112}. However, current assay needs improvement to make it more relevant for clinics and become a PoC technology. In this project, we will address three ways to optimize current FLEET platform, which are described below.

- Use *in silico* model to design ultra-specific probes for use in hybridization assay.
- Transfer existing FLEET bead-based assay to microfluidic chips.
- Expand FLEET platform to catch SNPs in microRNA.

By using *in silico* model for design of ultra-specific capture probes in FLEET assay, we hypothesize the melting temperature of LNA enriched capture probes can be predicted with high accuracy. In collaboration with Weber Group, we will apply the Peyrard-Bishop model to predict melting temperatures for a set of oncogene capture and linker probes with target. We will use the oncogenes *BRAF*, *EGFR* and *KRAS* that are known ctDNA biomarkers for NSCLC as proof of concept. To test the accuracy of the model, we will synthesize selected candidates based on the model calculations, measure T_m and compare with predicted T_m .

We aim to transfer existing FLEET hybridization bead-based assay to microfluidic chip. We propose to use poly(methyl methacrylate) (PMMA) as material for the microfluidic chip, and will immobilize ultra-specific $T_{10}C_{10}$ -capture probes, designed from the PB model, using UV light as

described in literature⁷². We aim to get same or improved sensitivity as the bead-assay using a spectrofluorometer as detection source for the chip.

Lastly, SNP profiling of miRNA has gain interest due to their link in inflammatory and cancerous diseases²⁴. Therefore, we aim to expand the FLEET platform to detect SNP mutations in miRNA. In previous work, miR128-2-3p was established as a potent biomarker for colitis and colon cancer in plasma samples¹¹². We want to use this biomarker as case of study to expand our platform and investigate SNP mutations in 3'-end to 3 adjacent bases of the miRNA128-2-3p. We will use the mesoscopic model to design capture probes for miRNA targets, enable discrimination between mutation and wild type. The method will be applied on colitis and colon cancer samples and compare with healthy samples, and verify using RT-qPCR.

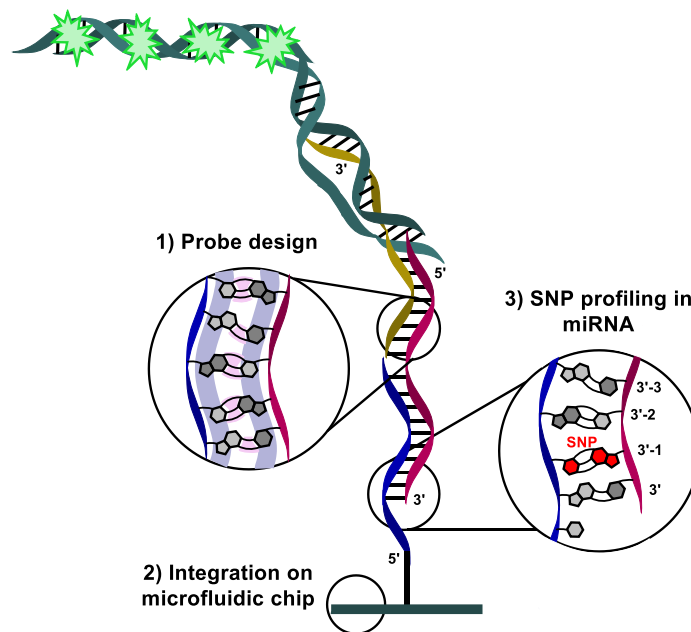


Figure 2.1: Optimizations of the FLEET assay have three focus areas. 1) Use Peyrard-Bishop model for probe design with LNA enrichments. 2) Integration of FLEET assay on microfluidic chips. 3) Extended FLEET assay for SNP profiling (3'-end to 3'-3) of clinical relevant miRNA biomarkers.

2.3 Aim of Project 2: Tandem RNA Synthesis

Related to oligonucleotide synthesis, the challenge of making multiple probes using solid-phase oligonucleotide synthesis for RNA therapy remain unsolved¹⁰⁴. Using solid-support, only one oligonucleotide is made at the time. Never the less, many gene therapies require at least two probes in the final formulation, such as *sense* and *antisense* strands in siRNA, heteroduplex ASO (DNA and RNA) and synergetic miRNA therapy^{42,87,113}. In this project, we hypothesize, that base labile cleavable moieties, that are orthogonal different from the TBDMS protection group on standard RNA-phosphoramidites, can be incorporated in *Tandem Oligonucleotide Synthesis* to make multiple RNA molecules (size~20nt) on the same support. Furthermore, we believe the yield of the outer oligonucleotide in the synthesized RNA-pool can be improved by incorporation of branching points to form dendritic structures.

For testing this, we will synthesize 2'/3'-OAcetyl Uridine phosphoramidites (normal and reverse) to form cleavable moieties that can be incorporated in the automated oligonucleotide synthesis between RNA sequences. As proof of concept, we will make pools of miRNA that are up- or down regulated in breast cancer. Furthermore, we will make linear- and branched pools to investigate the effect of branching effect on yield. For this, we will make standard curves from control miRNAs to calculate yield. As the branching point, we will synthesize a phosphoramidite having to DMTr-protected hydroxyl arms that can be incorporated in between the two cleavable nucleotides to make a branched structure. By using cleavable moieties suitable for *Tandem RNA Synthesis*, we believe we can make miRNA pools, which could have beneficial impact in production of therapeutic RNA in the future. Our approach is of relevance to the growing demand on synthetic RNA oligomers for nucleic acid research and technology.

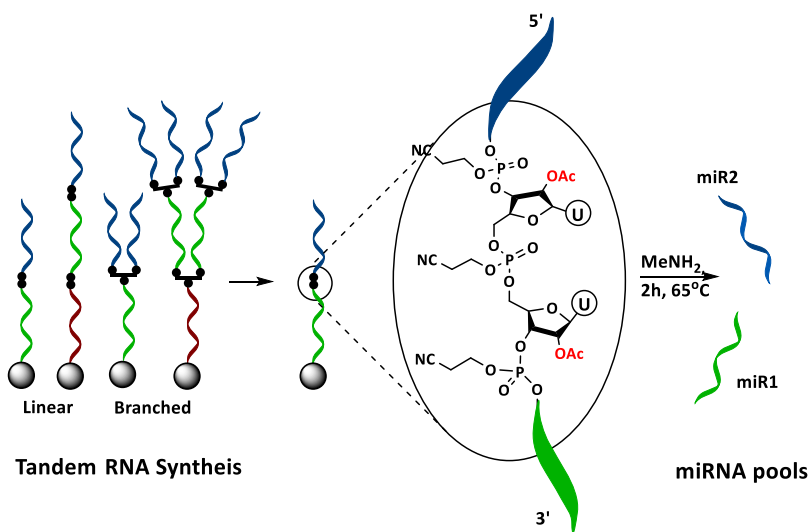


Figure 2.2: Tandem RNA Synthesis applied to synthesize microRNA pools in project 2.

2.4 Aim of Project 3: PCR-Free targeted single primer capture for ultra-deep sequencing

Targeted Sequencing methods have had great success for use in cancer diagnostics and precision medicine^{28,114}. They are very specific and sensitive, designed to only sequence the genomic regions of interest, thereby reducing cost and time compared to Next-Generation Sequencing. Targeted sequencing may not have relevance for PoC detection methods, since they are time consuming and complicated in sample preparation, however they still have significantly importance in detection *new* variants in the genome and biomarkers such as ctDNA and miRNA⁴⁶. One issue with current targeted methods is the use of PCR amplification in the sample preparation¹¹⁵. This can give PCR bias, meaning sequences are amplified unevenly and low abundance mutations can be lost¹¹⁶.

In the third and last project, we cooperated with Ji Research Group at Stanford University, to optimize an in-house targeted sequencing method called *Oligonucleotide-Selective Sequencing* (OS-Seq). The Ji Group developed the OS-Seq method such targets of interest were captured on the modified sequencing flow cell, limiting the complex sample preparation to only being library preparation (ref). Due to optimization in the Illumina platform making them more user-friendly, resulted in less freedom to modify the platform, and therefore moved to agarose beads. However, OS-Seq version 2 (OS-SeqV2) does not perform as well as previous version (OS-SeqV1), and needs to be optimized (OS-SeqV2.1). This project was part of my 6-month long external stay abroad.

Optimization of the *capture performance*, where captured target probes are extended on beads, could improve yield of generated library. Increased capture performance would result in improved sensitivity and more reads on the target region. Furthermore, we hypothesized capture performance could be improved by increasing the space between primers on the beads. For testing the hypothesis, we designed branching oligonucleotides with different spacing and incorporated them as primers in the OS-SeqV2.1. Previous examples of cone-shaped dendrimers on microarrays has shown improved capture of target and yield^{117,118}. The optimized OS-Seq method has potential to be an amplification-free targeted method, which could have great importance for rare mutation detections in the future.

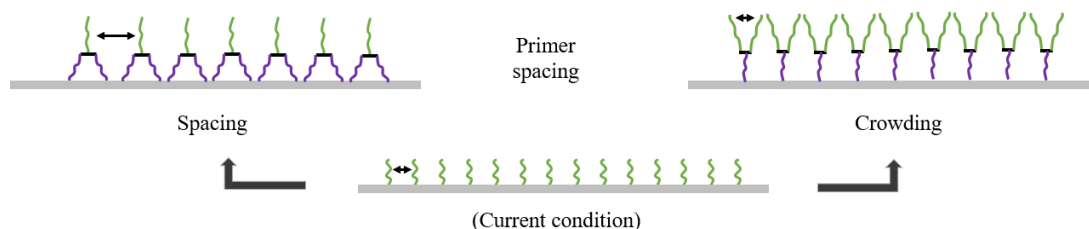


Figure 2.3: Branched strategy for primer spacing and crowding in project 3. Branched primers are used to optimize Oligonucleotide-Selective Sequencing on APEX beads.

3 A Diagnostic Platform for Early Detection of Cancer

3.1 Hybridization detection assay: FLEET

The FLEET platform was developed previously in the Astakhova group as an attempt to make a point-of-care platform for detection of nucleic acid cancer biomarkers, and is described in Uhd et al.¹¹¹. The word *fleet* originate from Old English meaning: “*moving very fast*”. It is an amplification-free hybridization assay on beads designed to detect ctDNA and miRNA from sera samples. In the method, target-specific biotin labeled capture probe is immobilized on streptavidin magnetic beads, from where target (miRNA or ctDNA from sample) is added to hybridize and make a complement match (**Figure 3.1**). In this assay a booster is applied which is a large DNA complex (calf thymus DNA, CTD, >100 kDA) to where many nucleic acid binding fluorophores can bind and thereby amplify the fluorescent signal. In order for the target:capture pair to bind the booster, a linker probe is needed which can bind the rest of target in one end, and booster in the other end. Linker and booster are mixed and added to target if present. After incubation, the beads are denatured to release the captured probe mix and transferred to a nucleic acid fluorophore solution (EvaGreen). If captures has recognized target, booster will be present, and the solution will give a fluorescent signal which can be analyzed with standard fluorometry and quantified with a standard curve.¹¹¹

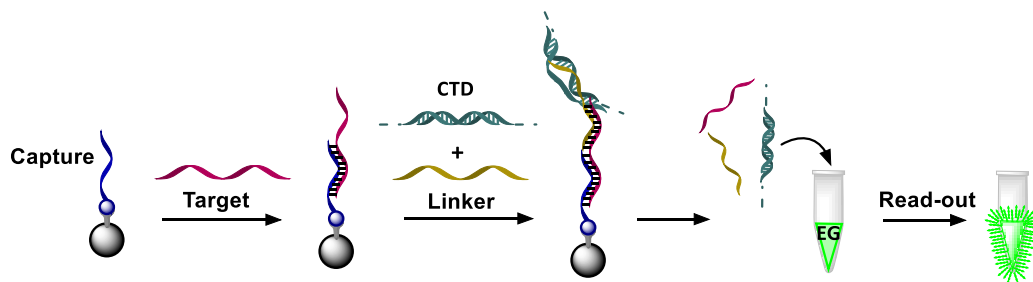


Figure 3.1: Principals of FLEET assay on beads. Biotin labeled capture probes designed for target probe (miRNA or ctDNA) are immobilized on streptavidin magnetic beads followed by addition of target, booster and linker to obtain a fluorescent read-out using nucleic acid dye¹¹¹.

Uhd et al. use the method to detect hsa-miR-223-3p and hsa-miR-486-5p with relevance for rheumatoid arthritis and cancer related mutations *BRAF* and *KRAS* of ctDNA¹¹¹. In previous work, we tested several miRNA in sera of human patients and in rodent model of colitis-associated colon cancer, using a tandem hybridization bead-based assay¹¹². We identified miR-128-2-3p as new potent biomarker for colon cancer progression. The current bead assay takes around 1.5 hour to run, and even with detectable photobleaching, the limit of detection is 10pM using standard fluorometry¹¹¹.

3.2 Peyrard-Bishop mesoscopic model for probe design

Locked nucleic acid (LNA) modifications are widely used in probe design to increase binding affinity and specificity to target probes^{119–123}. Nearest-Neighbor (NN) models are the most common models applied to predict thermodynamic properties of LNA-modified probes^{93,95–97}. Despite extensive use, NN models provide little insight into the intramolecular interactions and unable to indicate the physical sources of increased LNA stability. The Peyrard-Bishop (PB) mesoscopic model use the same melting temperature data as NN models, but have the potential to provide insights into the hydrogen bonds and stacking interactions (**Figure 3.2**)^{76,124,125}. In terms of hybridization temperature (T_m) predications, the PB-mesoscopic model has same capability as NN models, however it also allow stability analysis at base-pair level.

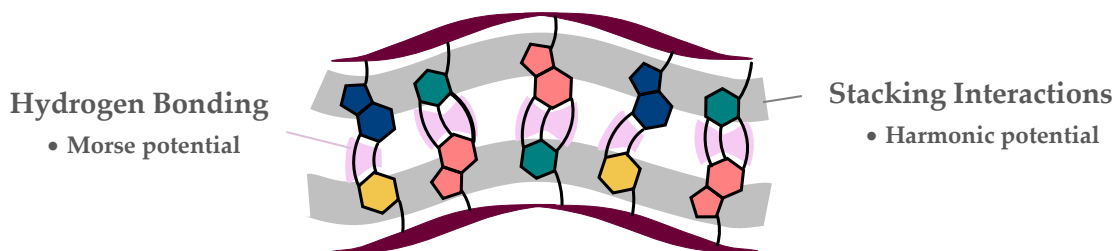


Figure 3.2: Intramolecular interaction in DNA+LNA:DNA helix. Peyrard-Bishop mesoscopic model describes hydrogen boning and stacking interaction in helix.

In brief, the Peyrard-Bishop model describes the DNA helix through a Hamiltonian, which contains a Morse potential describing the hydrogen bonds that connect each base-pair, and a harmonic potential describing the stacking interaction of adjacent base-pairs^{76,126–128}. By applying existing published data of single LNA modifications in DNA sequences and their experimental measured melting temperatures, it is possible to calculate a full parametrization of LNA modifications. This parameterization give insight into increased or decreased potentials due to

localized LNA insertions. Ferreira et al. found a substantial increase in Morse potentials of the LNA modified analogues compare to unmodified base-pairs (AT<CG≤A+T≤+AT<C+G<+CG) with biggest increase on +CG base-pair. This indicates stronger hydrogen bonding for LNA:DNA base-pair or new hydrogen bonds are formed giving the stabilizing effect⁹⁹. Stacking interactions of modified LNA was not increased compared to canonical unmodified, and conclude stacking interactions does not participate much to the enhanced stability of LNA⁹⁹.

Mesoscopic models like PB-model can be applied to predict melting temperatures of DNA:DNA, DNA:RNA hybrids and with insertions of LNA^{99,124,125,129,130}. The precision of the model depends on amount of existing dataset and sequences to be able to make calculations and predictions. The bigger dataset, the better performance of the model, and the more precise the prediction of melting temperature will be. The use of mesoscopic models can generate a better understanding of the thermodynamics and intramolecular interactions in modified sequence like DNA+LNA:DNA. Full mesoscopic parameterization of LNA modifications opens for the possibility to optimize the LNA inserts in oncogene probe design⁹⁹.

3.3 Project outline and goals

The FLEET assay has been developed in the Astakhova Group and applied to capture miRNA and ctDNA in sera samples comparable with current ways of detecting^{111,112}. However, existing platform can be improved to make it more relevant for clinics and closer to meet the point-of-care testing requirements.

3.3.1.1 *In silico* models to design ultra-specific probes for FLEET

By using *in silico* model for design of ultra-specific capture probes in FLEET assay, we hypothesize the melting temperature of LNA enriched capture probes can be predicted with high accuracy. In collaboration with Weber Group, we will apply the Peyrard-Bishop model to predict melting temperatures for a set of oncogene capture and linker probes with target. We will use the oncogenes *BRAF*, *EGFR* and *KRAS* that are known ctDNA biomarkers for NSCLC as proof of concept. To test the accuracy of the model, we will synthesize selected candidates based on the model calculations, measure T_m and compare with predicted T_m .

3.3.1.2 FLEET platform in microfluidic chips

Due to the recent development within microfluidics, we aim to transfer existing FLEET hybridization bead-based assay to microfluidic chip. We propose to use poly(methyl methacrylate) (PMMA) as material for the microfluidic chip, and will immobilize ultra-specific $T_{10}C_{10}$ -capture probes, designed from the mesoscopic model, using UV light as described in literature⁷². We aim to get same or improved sensitivity as the bead-assay using a spectrofluorometer as detection source for the chip.

3.3.1.3 Detection of miRNA and mutation variants

Lastly, SNP profiling of miRNA has gained interest due to their link in inflammatory and cancerous diseases²⁴. Therefore, we aim to expand the FLEET platform to detect SNP mutations in miRNA. In previous work, miR128-2-3p was established as a potent biomarker for colitis and colon cancer in plasma samples¹¹². We want to use this biomarker as case of study to expand our platform and investigate SNP mutations in 3'-end to 3 adjacent bases of the miRNA128-2-3p. We will use the mesoscopic model to design capture probes for miRNA targets, enable discrimination between mutation and wild type. The method will be applied on colitis and colon cancer samples and compare with healthy samples, and verify using RT-qPCR. This work is discussed in Chapter 4.

3.4 Results

3.4.1 Use of mesoscopic model to design ultra-specific probes

In this project, we collaborated with Isabela Ferreira, PhD from Professor Gerald Weber's group, Federal University of Minas Gerais in Brazil, who did the mathematical and physical calculations and T_m predictions with the mesoscopic model. We made the initial design of capture and linker probes, oligonucleotide synthesis and melting temperature study. It was published in Ferreira et al.⁹⁹. This will be a brief summary of results obtained from the study.

3.4.1.1 Probe design and synthesis

We designed seven different probes, for the microfluidic chip and T_m predictions. These were designed using publicly available DNA genome sequencing data from National Center for Biotechnology Information (NCBI) with GC content in the range of 38-55% and melting temperatures above 45°C according to the initial NN prediction for unmodified DNA in minimum salt buffer. This was done to ensure proper binding affinity of the capture and linker probes in the hybridization assay. The mutated oncogenes were assessed using dbSNP base: *BRAF V600E* (rs113488022), *KRAS G12D* (rs121913529), *KRAS G13D* (rs11244544) and *EGFR L858R* (rs121434568). Their respective NCBI codes are described in **Table S3.4**. Oligonucleotides were designed such the capture probes are complementary to the position of the corresponding gene bearing the mutation overlapping with 20-25nt, with the mutation positioned in the beginning of the sequence. The linker probe were complementary to the sequence downstream the gene with a gap of >20nt from capture probe. For enable immobilization of PMMA microfluidic chips, the capture probes were added the T₁₀C₁₀-tail in 5'-end and linker probes were added booster binding part in 5'-end.

Once the candidate probes for capture and linker were established, Izabela Ferreira made the calculations to find the optimized DNA+LNA:DNA parameters to calculate all possible configurations with one, two or three single LNA modifications for each selected probe. Between 834-2345 different LNA-modified sequences were calculated for each probe. The sequences with lowest number of LNA to the highest T_m , and highest discrimination between T_m full match and T_m mismatch for the corresponding oncogene mutation site were the ideal probes⁹⁹. Seven candidates were selected (4 capture and 3 linker) and were synthesized using solid-phase oligonucleotide synthesis. The capture and linker probes were all synthesized successfully with purity >90% after purification (**Table 3.1**). The purified oligonucleotides were applied for T_m study and preliminary studies of microfluidic FLEET assay.

Name	Sequence 5' → 3'	MW [g/mol]	MS obs. [m/z]	Purity
BRAFcap	T ₁₀ C ₁₀ - <u>ATCGAGAT</u> + <u>TTCT</u> + <u>CTGTAG</u> +CTA	12457.1	12455.9	90.6%
EGFRcap	T ₁₀ C ₁₀ -GAG+AAAAAGTTTCTCA+TGTA+CAGT	13431.7	13432.8	98.3%
KRAS12cap	T ₁₀ C ₁₀ -GCACTCTTGCCCTACCCA+ATC	11934.7	11933.8	99.6%
KRAS13cap	T ₁₀ C ₁₀ -GCACTCTTGCCCTA+CGCA+TTC	12007.8	12008.3	99.2%
BRAFlink	<u>GA</u> +TGG+GAATA+CCAGAC+CAC+CTGTTTTCA A+CTGTT+CAAA+CTGAT	13493.9	13493.0	97.8%
EGFRlink	<u>GA</u> +TGG+GAATA+CCAGAC+CAC+CTGTTTTTG AAGT+CA+CA+TTATATA	13518.9	13519.3	95.2%
KRASlink	<u>GA</u> +TGG+GAATA+CCAGAC+CAC+CTGTTTTTT G+TTG GAT+CATA+TTCGT	13823.0	13822.4	98.3%

Table 3.1. Oligonucleotide characterization. Fully underlined sequences are predicted target binding part of capture and linker probes, dashed underlined are booster binding part ($T_m = 69.7^\circ\text{C}$); + : LNA modification. Calculated mass, observed mass, and purity are shown in the table.

3.4.1.2 T_m Prediction and UV Measurements

To verify predicted T_m for synthesized probes, the melting temperature for each probe:target duplex was measured experimentally. Target probes of mutated oncogenes corresponded to the NCBI code sequence (shown in **Table S3.4**) and referred to as BRAF_{targ}, KRAS12_{targ}, KRAS13_{targ} and EGFR_{targ} in the rest of this chapter. Capture probes were paired with their corresponding target (0.5 μM of each) and likewise with linker probe to form the helices capture:target and linker:target pairs. The pairs were first annealed and then T_m determined as the maximum of the first derivative from resulting melting curve. The predicted and experimental melting temperatures (T_{pre} and T_{exp}) for each probe are shown in **Table 3.2**. The control (T_{ctrl}) refers to the unmodified DNA/DNA probe.

ID	Synthesized probe with LNA modification (5' → 3')	T_{ctrl}	T_{exp}	T_{pred}
BRAFcap-M1225	ATCGAGAT+TTCT+CTGTAG+CTA	59.6	64.5	64.9
BRAFlin-M416	CAA+CTGTT+CAAA+CTGAT	50.8	62.2	60.8
KRAS12cap-M1344	GCACTCTTGCCCTACCCA+ATC	64.1	64.7	65.4
KRAS13cap-M1304	GCACTCTTGCCCTA+CGCA+TTC	64.2	68.9	68.8
KRASlin-M616	TGAAGT+CA+CA+TTATATA	48.2	55.0	58.2
EGFRcap-M943	GAG+AAAAAGTTTCTCA+TGTA+CAGT	57.4	61.2	61.8
EGFRlin-M478	TTG+TTGGAT+CATA+TTCGT	54.5	61.5	61.5
$\langle T_{\text{exp}} - T_{\text{pred}} \rangle = 0.91^\circ\text{C}$				

Table 3.2. Measured T_{exp} , Control T_{ctrl} , and Predicted Temperatures T_{pred} for the Selected Sequences from the Probe Prediction. The probe identification (ID) refers to the candidate probe list in Supporting Information Tables S5–S11 in Ferreira et al.⁹⁹ The average deviation $\langle T_{\text{exp}} - T_{\text{pred}} \rangle$ was calculated between the available predicted and measured temperatures. Table reprinted from Ferreira et al.⁹⁹

From the table, the measured melting temperatures (T_{exp}) differentiate from the predicted melting temperatures (T_{pre}) with 0.91°C in average. This shows a good agreement between predicted and measured, and confirm the Peyrard-Bishop model can be used to design specific LNA enhanced DNA probes with good accuracy. Only few discrepancies was observed such as KRASlink predicted 3°C more than measured (**Table 3.2**), indicating the model can be improved. This could be done using more data in the parameterization of LNA, such the prediction gets more precise. However, *in silico* models are limited by the available data and it takes time to produce more data to optimize the model. Moreover, the PB model is assuming a flat complex structure (2D) for the nucleic acid hybrid. In some cases this assumption might be acceptable such as in DNA+LNA:DNA predictions, where mismatches are not tolerated much. However, for RNA targets such as miRNA, this could be a problem since they fold in 3D structure. This could be something to look at in the future for PB model in probe design.

3.4.2 FLEET using Microfluidic Chips

For clinical relevance, we wanted to transfer existing FLEET assay from streptavidin magnetic beads to a PMMA microfluidic chip. Despite exiting and important challenge, it was not an easy task. The assay consist of many steps, and many things needed to be optimized. The initial idea of this project, was to make a microfluidic chip capable of analyzing a panel of different biomarkers for testing blood samples for ctDNA or miRNA (**Figure 3.6**). The chip should be sensitive and selective to be reliable. Secondly, it should be easy to use, easy to read, cheap in materials but have good optical properties. Therefore, we decided to use poly(methyl methacrylate) (PMMA) as material with the TC-tag capture probes designed using the mesoscopic model^{72,99}.

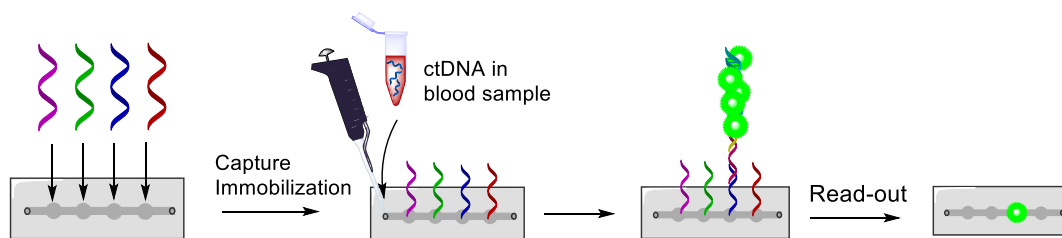


Figure 3.6: Goal of microfluidic FLEET platform, to have a panel of capture probes, enable screening for several NSCLC biomarkers at the same time.

We started to make a prototype of a chip with help from Professor Martin Duvfa, DTU Health Technology, who had the equipment, materials and experience to fabricate chips. MSc student Asmund Kjellegaard Ottesen, DTU Physics, help with the initial design, fabrication and test of the microfluidic chip. BSc students Anna Kyung Jakobsen and Marie-Lousie Wibom, DTU Chemistry, made the second batch of fabricated microfluidic chips as part of their bachelor thesis. BSc student Cecilie Schiøth Krüger-Jensen did the dye and photobleaching study described in Chapter 4. Two different prototypes were designed and fabricated; 1-chamber and 4-chamber chips were made, however only 1-chamber chip were applied due to issues and lack of time. The fabrication and design of PMMA chips are described in experimental section. In brief, the microfluidic chips were designed and fabricated in three layers and bonded together to make a microfluidic chip with inlet on top, closed reaction chamber and outlet hole in the bottom (**Figure 3.7**, right).

3.4.2.1 Initial Tests of Chip in Spectrofluorometer

Initially, the chip was tested for signal in the Edinburgh Spectrofluorometer FS5 by EvaGreen and booster solution in one-chamber chip. The limit of detection was estimated to be between 6.4pM to 32pM by making a dilution row (**Figure 3.7**, left). A blank and empty was also made. LOD for the bead assay was 10pM with standard fluorometry, so this method operates in the same range.

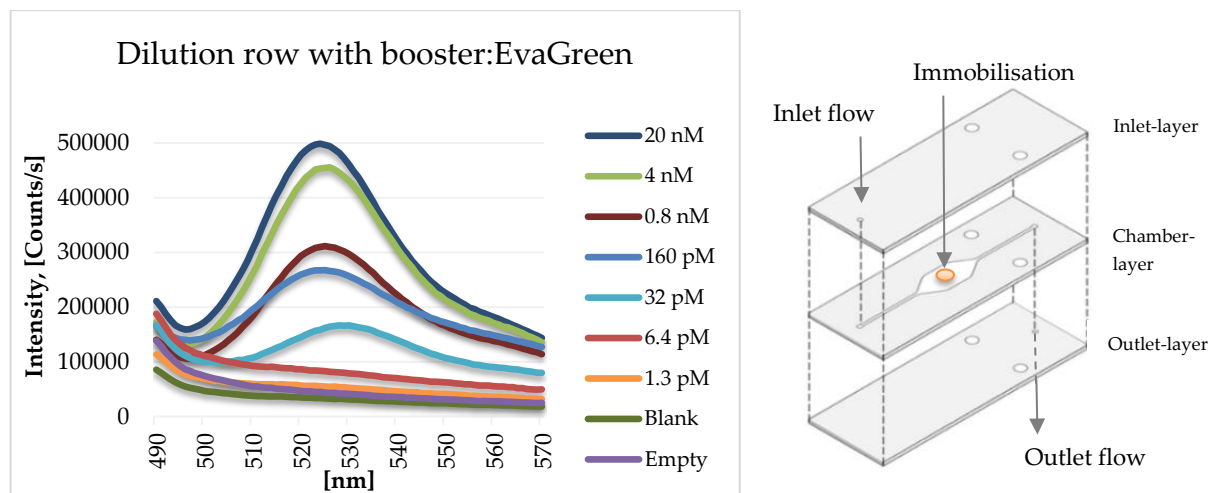


Figure 3.7: Dilution row of EG:booster (left). One-chamber chip design shown with 3 layers (right)

3.4.2.2 Preliminary test of microfluidic FLEET with BRAF target

To incorporate the FLEET platform into microfluidic PMMA chips, we used *BRAF* oncogene probes (BRAFCap, BRAFTarg, BRAFLink) as case of study, to see if we could run the whole assay and get a useful signal. The microfluidic assay is described in brief below.

Method. The microfluidic PMMA chips were first prepared and immobilized with T₁₀C₁₀-probe (2 μ l, 50 μ M) using UV light (30 sec) as described in literature⁷². The chip was closed, and unbound probe was washed away with PBS buffer using micropipette. Target-probe solution (40 μ l, 50 nM) was added to the chip to react for 20 min, and then washed 4 times with heated buffer (45°C). Linker-probe was added to react in 20 min and washed 4 times with heated buffer (45°C). Next, booster solution (200 nM) was added to react in 20 min, and then washed 4 times. Finally, EvaGreen solution was added to the chip and the fluorescence is measured using Edinburgh Spectrofluorometer FS5 where the chip was mounted and emission measured (490-570 nm).

The assay was prepared in one-chamber chips (Figure 3.7, right) and washing buffer (1xPBS) was varied in temperature (25°C or 45°C) with addition of Tween (0%, 0.02%, 0.2%). It was seen from Figure 3.8 (upper) that chips washed with heated buffer at 45°C (chip 4-6) in general are giving a higher signal. Furthermore, the chips washed with 0.2% tween in buffer gives higher signal than the 0.02% and 0%. The best washing conditions for the pilot assay was set to be 1xPBS + 0.2% Tween at 45 °C. Unfortunately, it was also seen that the blank actually gave a signal, which questions the reliability of our results. We proposed the signal could arise from booster left in chip after 4 washes. We washed the blank 3 times more, which improved the signal, to be close to negative curve. It was concluded from the first pilot assay that at least 7 washes are needed after booster step to obtain a reliable signal.

The experiment was repeated with varying incubation temperatures for the hybridization steps (25°C or 37°C) and tween concentrations (0.2%, 0.5% and 1%) for washing buffers (1xPBS, 45°C) and with 7xwashes after booster addition. The results are seen in **Figure 3.8** (lower): Blank did not give any signal, which makes the results reliable. However, the signals were in general very low, and much lower than previous experiment. Furthermore, it was difficult to see any trend within the parameters, and not consistent. Chip 3 gave the strongest signal (1% tween with 37°C incubation), therefore these conditions were chosen for the last preliminary study regarding wild type discrimination.

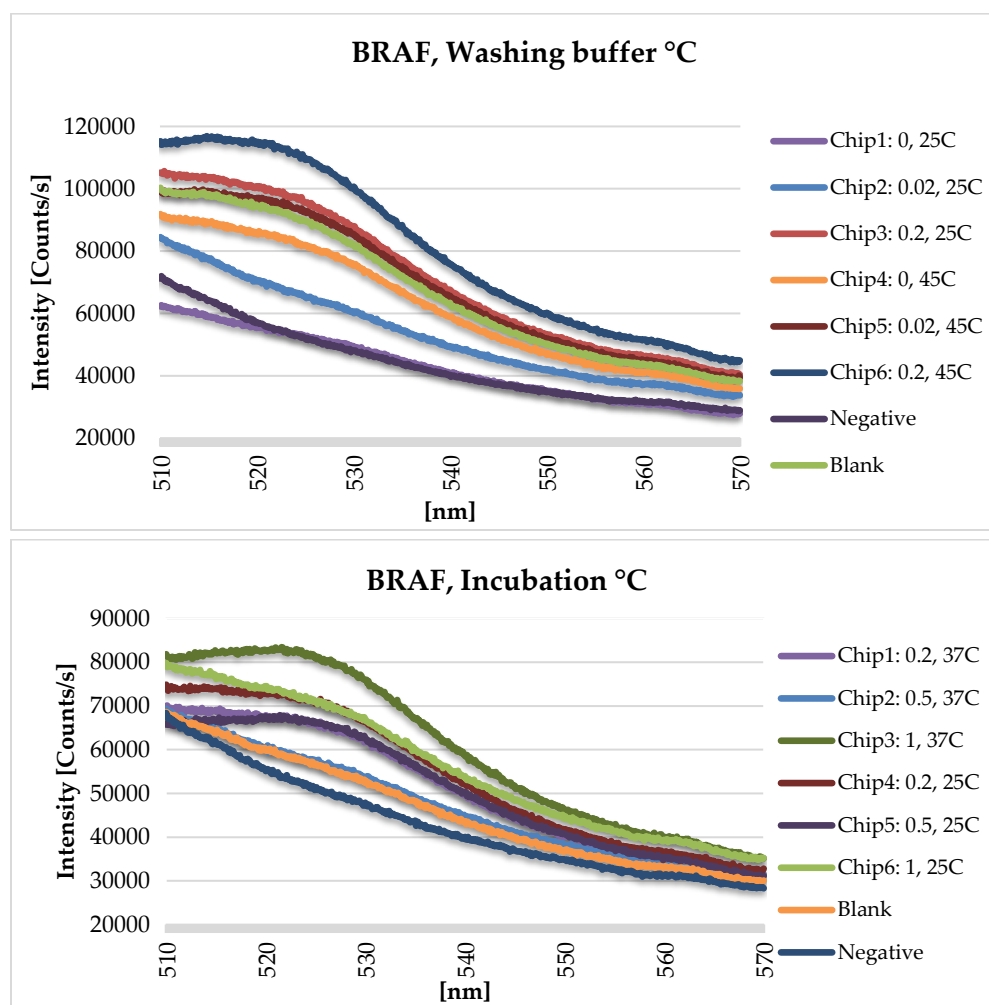


Figure 3.8: Fluorescent readout from FLEET chip assay capturing BRAF target in one-chamber chip. (Upper) Assay with different washing buffer conditions (Tween% and temperature). (Lower) Assay with different incubation temperature and tween% in washing buffer.

3.4.2.3 Mutation vs. wild type discrimination

In the last test, we investigated the ability of capture probe to discriminate between SNP-target (mut) and wild type (wt). This is important in order to ensure high specificity of the FLEET performance in microfluidic chips and to avoid false positive. The FLEET assay was made in duplicates (mean is show below), with optimized conditions examined from previous tests (washing buffer: 1xPBS+1%Tween, 45°C, incubation 37°C, and 7xwashes after booster step). Target samples (50nM) were prepared in ratios: 0-100% of BRAFtarget:BRAFwt. Target mutation and wild type were mixed to look at potential effects when different concentrations are present at the same time. To verify the signals, a scramble probe and a blank were prepared. The sequences of BRAFtarg, BRAFwt and scramble are shown in **Table S3.2**. Results of experiment is shown in **Figure 3.9** as a bar plot.

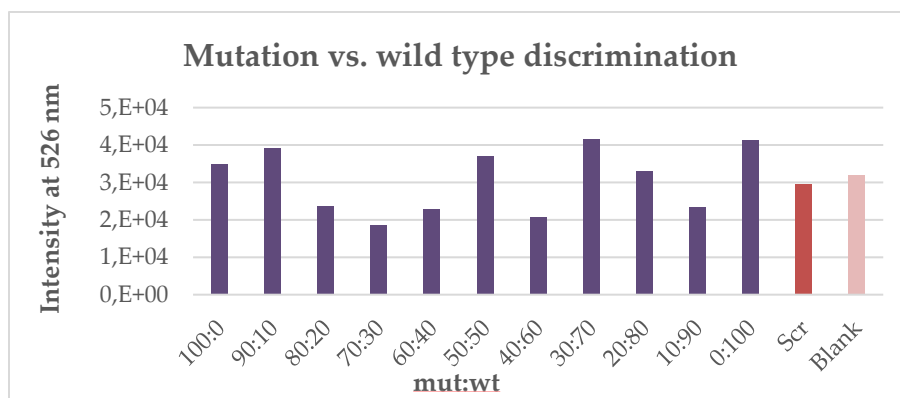


Figure 3.9: Mutation vs. wild type discrimination using FLEET assay. Intensity measured at maximum at 526nm for ratio between mut:wt. Scr: scramble and blank were also measured.

We expected to see that all mut:wt mixtures would give a signal, ideally strongest at target mutation and lowest at wt. Despite it looks like mutation and wild type binds to the assay, scramble and blank also show emission which is problematic since both should be zero. This is indicating that either booster remain in the chamber after 7 washes, or capture probe is not immobilized, such the only signal comes from trapped booster. Due to issues with low intensity and difficulties with consistence in measured samples, we decided to go several steps back and ensure prober immobilization using fluorescence microscopy.

3.4.2.4 Immobilization study

We wanted to test each step in the assay with fluorescent microscope, however due to limited time with the fluorescent microscope, it was decided to make a simple setup with a fluorescent probe to test immobilization on PMMA. For this experiment, we made KAS246c: 5'-TTTTTTTTT-TTCCCCCCCCC-[AlexaFlour488]-3' by synthesizing the 3'-Amino-Modified T₁₀C₁₀ probe which

was conjugated to AlexaFlour488-NHS ester post synthesis (Alexa Flour 488, ex/em: 496/519 nm). The conjugation was confirmed by MALDI-TOF, however complete conversion was not obtained, but it was decided to use it for immobilization test. The probe was prepared in 50 μ M conc. and 2 μ L immobilized with UV light in one-chamber chips. We tested one chip before washing and one after 3 x washing (**Figure 3.10**). It was seen from the microscope pictures that after 3 washes some probe was left on the chip indicating immobilization has happen between probe and PMMA to some extent. However, signal to noise ratio is still too low to detect with Edinburgh Spectrofluorometer, so it was decided to try to shift to another material than PMMA.

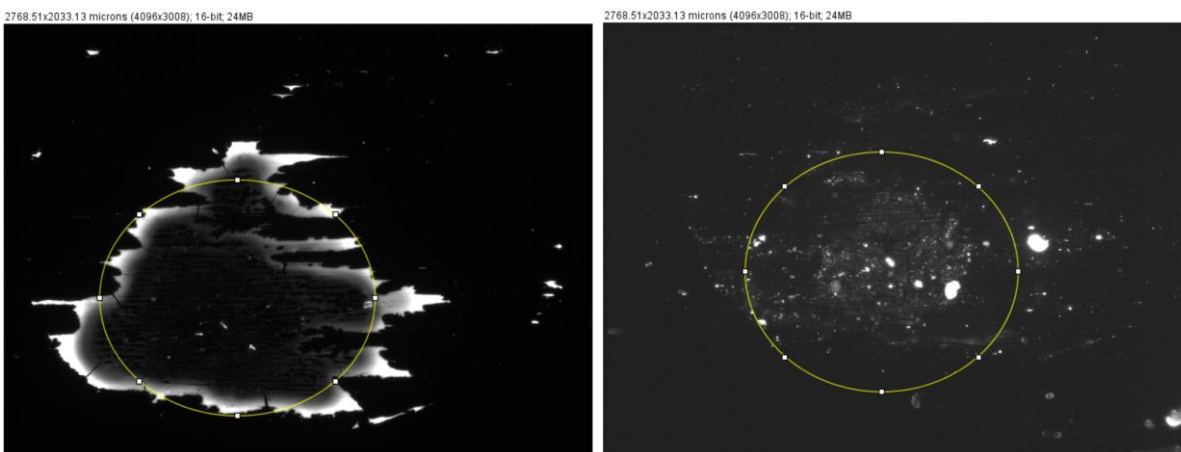


Figure 3.10: Fluorescence microscopy pictures of KAS246c immobilized in one-chamber PMMA chip before wash (left) and after 3 washes (right).

Instead of immobilizing the DNA via UV light and photochemistry, PMMA can be functionalized to have aldehyde on the surface such that amino-modified DNA probes can be immobilized⁷⁴. However, most arrays and microfluidic chips are glass slides modified with some functional group⁷⁰. TC-tagged DNA probes can be immobilized on glass slides or amino-modified DNA probes on aldehyde-glass slides can be used⁷⁵. Due to limited time, and issues with providing glass slides, we did not try FLEET assay using glass slides. However, this is something we would like to try in the future.

Furthermore, it would be good to use established detection methods for slides and microarray such as ScanArrayLite or fluorescence microscope to ensure good detection in all steps^{75,131}. Another issue with current microfluidic FLEET method is that all reagents are added manually with micropipette and that outlet reagents are removed with capillary force. It is essential to optimize chip material, design of chip, and potential to make the process automated, to ensure high throughput and reduce personal errors to maintain consistency in the assay.

3.5 Conclusion and Future work

In summary, we have tried to optimize current FLEET procedure established in Astakhova group for detection of miRNA and ctDNA in sera samples. For that, we have applied an *in silico* model to predict melting temperatures for capture and linker probes modified with LNA inserts. From the predictions, sets of oncogenic LNA enhanced capture and linker probes were synthesized using solid-phase synthesis. The melting temperatures of the synthesized probes with their respective targets were measured and showed accuracy within 1°C from the predicted. *In silico* models for prediction of melting temperature and probe design has great potential. Further optimizations of the PB model for broader use and application could be implementing 3D structures of target as a parameter for the model. This is relevant for RNA targets that fold in secondary structures. Moreover, we applied the PB model to look at DNA+LNA:DNA hybrids, but it could also be interesting to look at other nucleic analogues or conjugates to expand the PB model as prediction tool. However, this would require data set of experimental measured hybrids with analogues or conjugates.

We aimed to implement the FLEET assay in microfluidic PMMA chips using TC-tagged capture probes immobilized with UV light⁷². For this purpose, one-chamber and 4-chamber chips were designed and fabricated in three layers and bonded together to make a microfluidic chip with inlet on top, closed reaction chamber and outlet hole in the bottom. The limit of detection was estimated with Booster:EvaGreen solution to be 6.4pM – 32pM using a spectrofluorometer. The assay was tested with BRAF-probes designed using the Peyrard-Bishop model as proof of concept, showing a detectable signal. However, even after several optimizations with extra washes heated buffers and incubation time, signal remained very low and inconsistent close to blank values. Immobilization on PMMA was studied using a fluorescent TC-tagged probe visualized with fluoresces microscopy. This showed immobilization on PMMA to some extent; however, it was decided to try the FLEET method in more established materials and detection methods such as glass slides in the future.

Implementing the FLEET assay in microfluidic chips in the course of making a PoC platform is an exciting and challenging interdisciplinary task. However, many optimizations steps are needed to reach a final product. A future strategy for this could be to ensure chemistry is working on established platforms similar to microarrays before moving to cheaper options.

3.6 Experimentals

3.6.1 Oligonucleotides

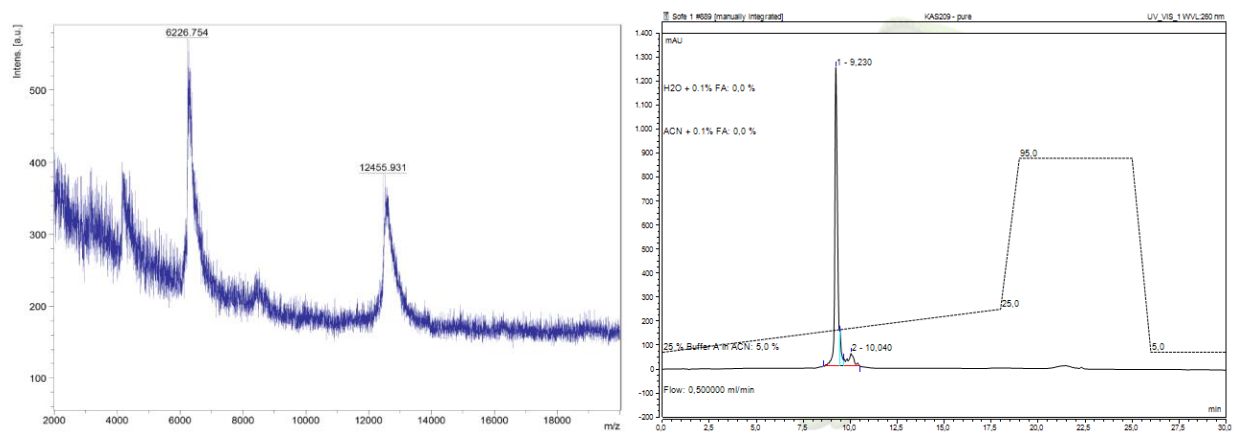
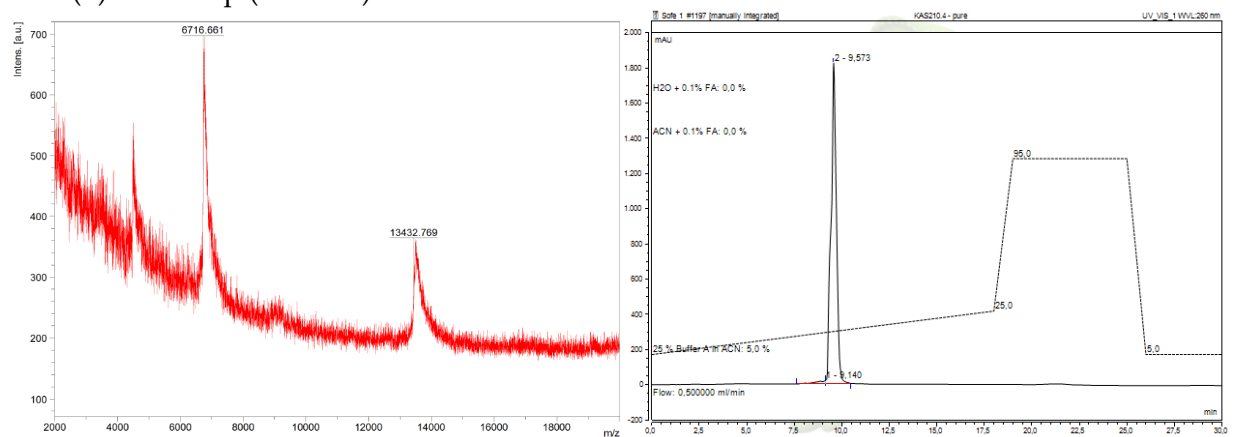
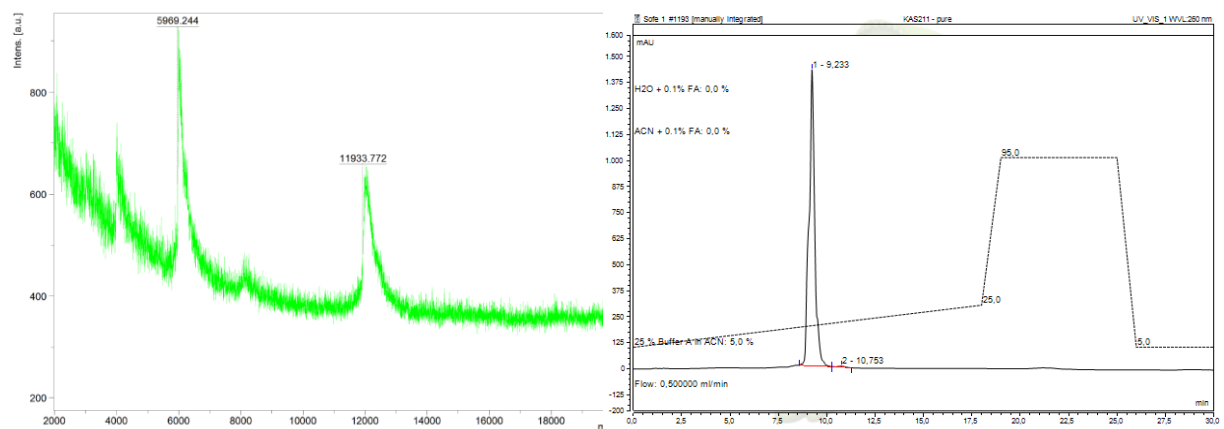
3.6.1.1 Oligonucleotide Synthesis

The LNA/DNA capture and linkers were synthesized in 1μmol scale using solid support (CPG 1000Å, Sigma Aldrich) on Expedite Nucleic Acid Systems or Biosset DNA/RNA Synthesizer. Reagents purchased from Sigma Aldrich: TCA Deblock, DCI activator 0.25M, Oxidizer 0.02M, CapA and CapB. Commercial phosphoramidites from Sigma Aldrich (dA(Bz), dC(Bz), dG(ib), dT) were all prepared in 0.07M solutions using dry acetonitrile (MeCN). LNA phosphoramidites (indicated with '+' in table below) from Qiagen (LNA-A(bz), LNA-mC(Bz), LNA-G(dmf), LNA-T) dissolved in 0.07M in dry MeCN (+C: 50:50 DCM:MeCN). The oligonucleotides were cleaved from solid support using 28-30% ammonia solution from Sigma Aldrich at 55°C, 12h. The identity of the oligonucleotides were established by mass spectrometry (MS) using an Autoflex speed MALDI-TOF mass spectrometer (Bruker Daltonics, Hamburg, Germany). Samples were co-spotted with 3-Hydroxypicolinic acid as matrix on a MTP AnchorChip target plate for the analysis and run at linear positive detection mode. The obtained mass spectra were recorded by the flexControl 3.4 (Bruker Daltonics, Germany) software. The oligonucleotide was purified on an Ultimate 3000 UHPLC (Dionex, Sunnyvale, CA, USA) using a DNA-Pac RP (Thermo Fisher Scientific, Waltham, MA, USA) column (4μm, 3.0 × 100 mm²) with a gradient of 5–25% buffer B in A over 30 min at 60°C. Buffer A: 0.05M TEAA; Buffer B: 25% A in MeCN. Peaks were monitored at 260nm. HPLC and MALDI spectra are shown below. A stock of each pure oligonucleotide was prepared in concentration of 100μM in MQ.

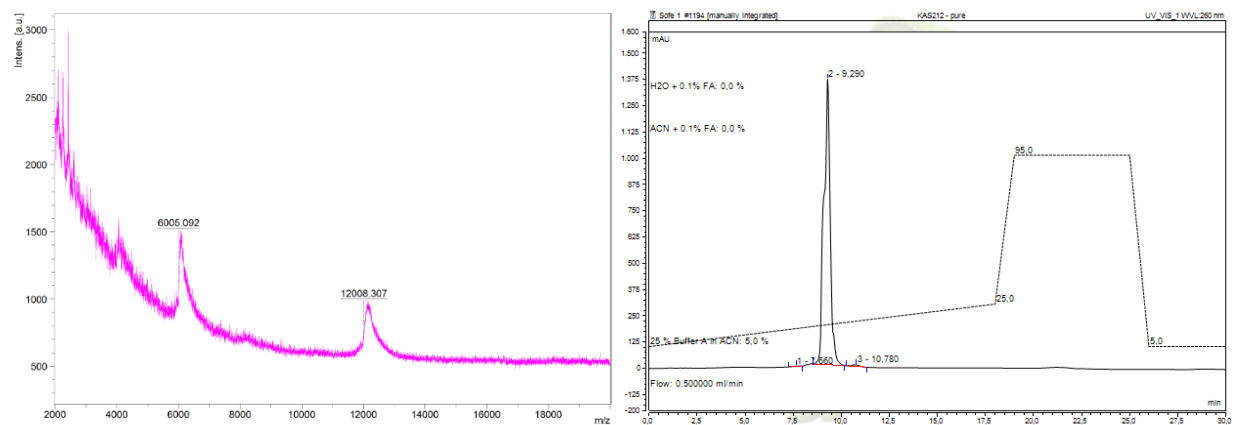
3.6.1.2 Characterization

Name	Sequence 5' → 3'	MW [g/mol]	MS obs. [m/z]	Purity
BRAFcap (KAS209)	TTTTTTTTTCCCCCCCCCATCGAGAT+TTCT+CTGTAG+CTA	12457.1	12455.93	90.6%
EGFRcap (KAS210)	TTTTTTTTTCCCCCCCCCGAG+AAAAAGTTTCTCA+TGTA+CAGT	13431.7	13432.77	98.3%
KRAS12cap (KAS211)	TTTTTTTTTCCCCCCCCCGCACTCTTGCCTACCCA+ATC	11934.7	11933.77	99.6%
KRAS13cap (KAS212)	TTTTTTTTTCCCCCCCCCGCACTCTTGCCTA+CGCA+TTC	12007.8	12008.31	99.2%
BRAFlink (KAS213)	GA+TGG+GAATA+CCAGAC+CAC+CTGTTTTCAA+CTGTT+CAAA+CTGAT	13493.9	13492.97	97.8%
EGFRlink (KAS214)	GA+TGG+GAATA+CCAGAC+CAC+CTGTTTTGAAGT+CA+CA+TTATATA	13518.9	13519.30	95.2%
KRASlink (KAS215)	GA+TGG+GAATA+CCAGAC+CAC+CTGTTTTTG+TTG GAT+CATA+TTCGT	13823.0	13822.41	98.3%
KAS246	TTTTTTTTTCCCCCCCCC-X	6080.98	6089.93	90.0%
KAS246-c	TTTTTTTTTCCCCCCCCC-[AlexaFluor488]	6604.91	6603.52	-

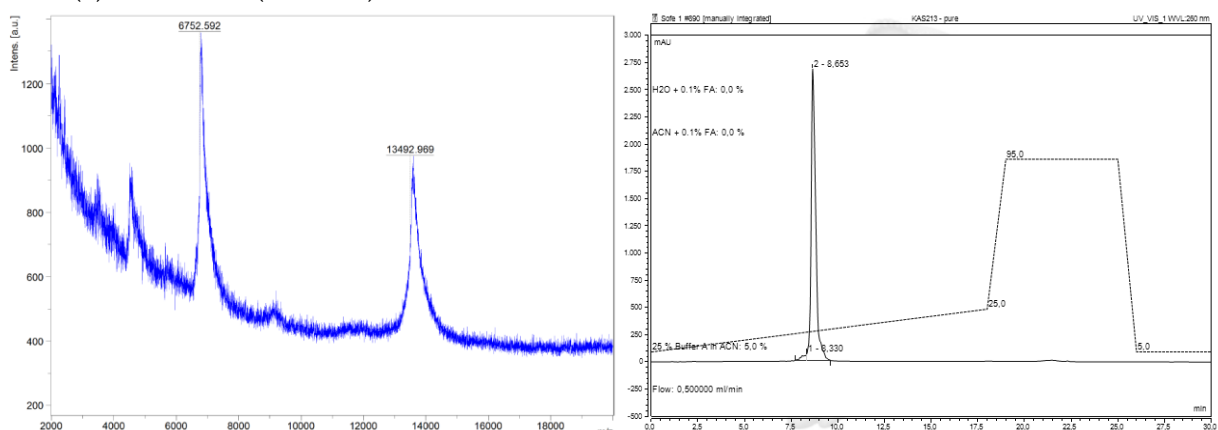
Table S3.1. Oligonucleotides synthesized. X: 3'-Amino-Modifier CPG.

(a) BRAFc_{ap} (KAS209)(b) EGFRc_{ap} (KAS210)(c) KRAS12c_{ap} (KAS211)

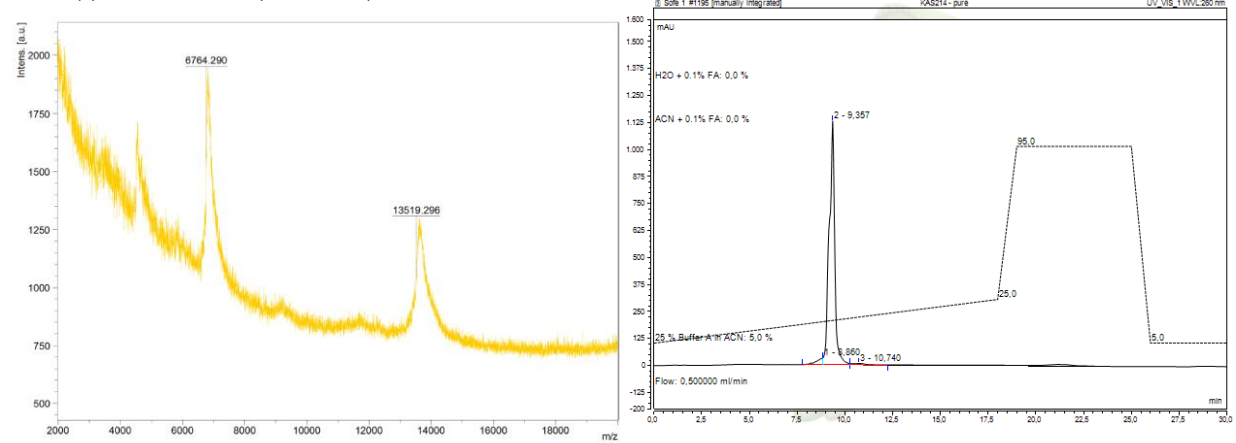
(d) KRAS13cap (KAS212)



(e) BRAFlink (KAS213)



(f) EGFRlink (KAS214)



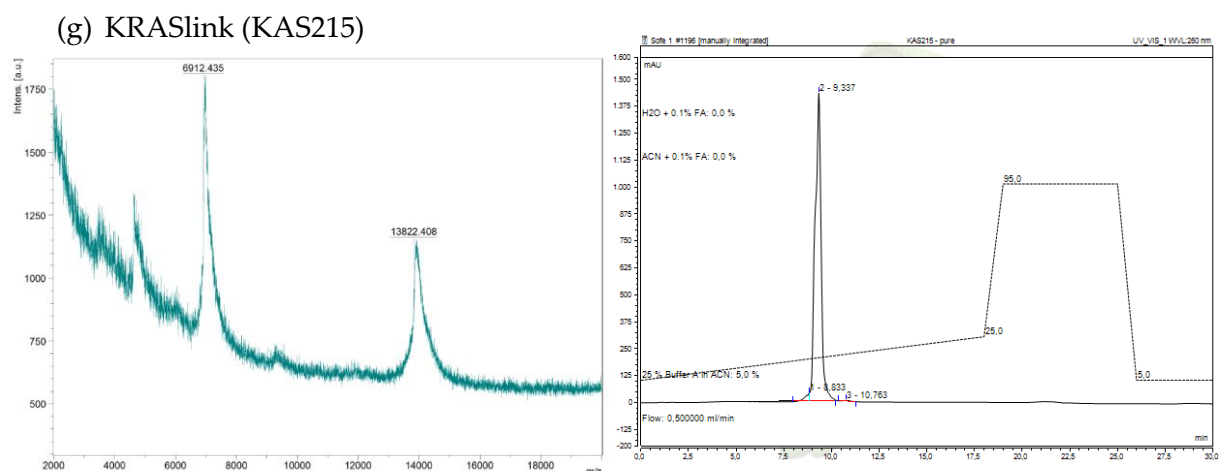


Figure S3.1: MALDI-TOF spectra (left) and HPLC trace (right) of pure oligonucleotides; (a) BRAFc_{ap}; (b) EGFRc_{ap}; (c) KRAS12c_{ap}; (d) KRAS13c_{ap}; (e) BRAFl_{ink}; (f) EGFRl_{ink}; (g) KRASl_{ink}. Data reprinted from Ferreira et al.⁹⁹

3.6.1.3 NHS Coupling of Alexa488 to KAS246 for Immobilization

To an Eppendorf tube, crude 3'-Amino-Modified oligo (KAS246, 20nmol, 8.2μL) was added followed by addition of 30μL 0.1M bicarbonate buffer. To the mixture 41.8μL DMSO was added followed by 20μL (10 eq., 200nmol) NHS activated fluorophore (Alexa Fluor 488 NHS Ester, Invitrogen) freshly prepared from stock (1mg/180μL DMSO) to get a final volume on 100μL. The mixture was vortexed, covered with foil and left to react at rt. in the dark o/n. Excess fluorophore was removed using NAP-5 Columns Sephadex from GE Healthcare, eluted with MQ water. The fluorophore conjugated oligonucleotide (KAS246c) was characterized by HPLC and MALDI-TOF to confirm conjugation.

3.6.1.4 Purchased Oligonucleotides for Microfluidic

We designed targets equivalent to the genome site of the oncogenes *BRAF V600E*, *EGFR L858R*, *KRAS G12D* and *KRAS G13D* from gene bank having one part complementary to capture design and another part complementary to linker sequence. DNA target sequences were purchased at IDT along with wild type and scramble sequences.

Name	Target sequences 5' → 3'
BRAF _{targ}	5'-TAGCTACAGAGAAATCTCGAT-TTTT-ATCAGTTTGAACAGTTG-3'
BRAF _w t	5'-TAGCTACAGTGAATCTCGAT-TTTT-ATCAGTTTGAACAGTTG-3'
BRAF _{scr}	5'-ATTGATTCCATGTTCTGTATAACGAGATGTCGATATACTGAT-3'
EGFR _{targ}	5'-GGGCGGGCCAAACTGCTG-TTTT-AGAAAGAATACCATGCAGAA-3'
EGFR _w t	5'-GGGCTGGCCAAACTGCTG-TTTT-AGAAAGAATACCATGCAGAA-3'

KRAS12targ	5'- <u>G</u> ATGGCGTAGGCAAGAGTGC-TTTT-CAGCTAATTCAGAATCATT-3'
KRAS13targ	5'-GGT <u>G</u> ACGTAGGCAAGAGTGC-TTTT-CAGCTAATTCAGAATCATT-3'
KRASwt	5'-GGTGGCGTAGGCAAGAGTGC-TTTT-CAGCTAATTCAGAATCATT-3

Table S3.2: Purchased oligonucleotides from IDT. wt: wild type; scr:scramble control; Underlined 5'-end is the capturing binding sequence and dashed underlined 3'-end is the linker binding sequences.

3.6.2 FLEET assay on chip

3.6.2.1 Chip Design and Fabrication

The microfluidic chips were made of poly(methyl methacrylate) (PMMA) bonded together with PSA (Pressure-sensitive adhesive) constructed with an inlet hole in first layer, main chamber in second and outlet hole in third layer and measure $1.7 \times 14 \times 41 \text{ mm}^3$ (**Figure S3.2**). The microfluidic chips were fabricated using laser ablation in 3-layers of 0.5mm PMMA bonded with 0.1mm PSA engraved with CO₂-laser (Epilog Mini 18, 30-watt, Laser cutter). CorelDraw software was applied for drawings for the laser cutter. Prior to cutting, the thin layer of protection plastic was removed from the PMMA surface and the PMMA piece was placed in the epilog on a 1cm plastic piece to reach the right height for cutting. The outlet - and inlet layer were engraved from drawings with the settings: 2xscans, power 30%, speed 20%, 600 DPI. The middle-layer was added 0.1mm PSA on both sides and engraved with 3xscans, 20% power, 20% speed and 600 DPI. Before bonding, the thin film of both outlet – and inlet layer was removed. The 3 layers of PMMA were combined, aligned and bonded together (20kN for 1 min), and broken into smaller chips with individual size $1.7 \times 14 \times 41 \text{ mm}^3$ and chamber volume of $40 \mu\text{l}$ as shown in **Figure S3.2**.

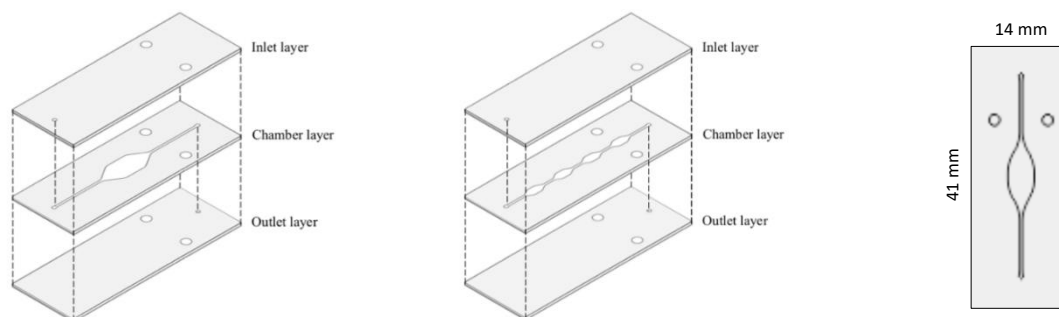


Figure S3.2: One-chamber (left) and 4-chamber (middle) microfluidic chips. Individual layer in the chip aligned. Collected one-chamber chip metrics; $1.7 \times 14 \times 41 \text{ mm}^3$ and chamber volume of $40 \mu\text{l}$.

Chip specifications	One-chamber chip	4-chamber chip
Area [mm ²]	40 × 20 × 2	40 × 20 × 2
Thickness [mm]	1.7	1.7
Chamber size [mm ²]	35	1
Number of chambers	1	4
Volume [μL]	40	25
Excitation λ [nm]	490 – 500	490 – 500
Emission λ [nm]	515 – 530	515 – 530

Table S3.3: Chip specifications of one – and 4-chamber chip.

3.6.2.2 Assay Procedure in Chip

Step 1: Immobilization. The immobilization of probes to PMMA chip was done using published procedure⁷². To initiate the cross-link between T₁₀C₁₀-capture probe and PMMA, the synthesized capture sequences were prepared individually in a stock of 75mM in 1xPBS containing 0.01% Triton to a final conc. of 50 μM. The chip were rinsed with 70% ethanol and dried for 10min (air blow gun was used to remove droplets). 2μL of the DNA capture solution were spotted in the middle of the chamber with pipette and was dried for 10 min. The UV lamp (DYMAX EC 5000 w. p/n 36970 bulb (225 mW/cm²)) was started 5min prior to use. Each chip with spotted capture probe was exposed for 30s with a distance of 12cm from the UV lamp⁷². Post exposure, the open chip with immobilized probe was bonded together with a top layer to close the chip. The chip was assembled with bonding pressure (1.0x10⁵ kN/m²) due to pressure sensitive adhesives. One chamber volume (40μl) of buffer solution (1xPBS) and left for 10 min to remove unbound probes. The solution was removed by capillary force and wash repeated once more. The chips were stored at -20°C if not applied the same day.

Step 2: Hybridization to target. To the immobilized capture probe, one chamber volume (40μl) of target solution (50nM in 1xPBS) was added to the chip and left to react for 20min at 37°C. The solution was removed with capillary force and washed with heated buffer (1xPBS + 0.02% Tween, 45°C) 4 times in total for each chip.

Step 3: Hybridization to linker. To hybridize the linker, one chamber volume of linker solution (50nM in 1xPBS prepared from stock solution) was added to the chip and left to react for 20min at 37°C. The solution was removed with capillary force and washed with heated buffer (1xPBS + 0.02% Tween, 45°C) 4 times in total for each chip.

Step 4: Booster. A solution of 200nM booster in MQ was prepared from a stock 2mg/ml. One chamber volume was added to the chip and left to react for 20min at 37°C with gentle shaking. The solution was removed with capillary force and each chip washed 7 times with 1xPBS buffer.

Step 5: EG and Fluorescent detection. Lastly a fluorophore solution (EvaGreen 5 μ L to 100 μ L MQ) was prepared and 40 μ L added to each chip. The fluoresces of each chip was analyzed using Spectrofluorometer FS5 from Edinburgh Instruments with a SC-10 Front-Face Sample Holder module, excitation at 470nm and spectra recorded from 510nm – 580nm. Before each a series of measurement, the equipment was calibrated with a positive booster control (32nM in MQ + EG) and negative control (0nM + EG) to ensure maximum counts/s. for the chip.

3.6.3 T_m Measurements

The melting temperature studies were performed on a DU800 UV/VIS spectrophotometer equipped with a Beckman Coulter Performance Temperature Controller. Complementary strands (0.5 μ M of each strand), in 1xPBS were mixed, denatured for 10 min at 90°C, and subsequently cooled to 15°C, the temperature at which the experiment was started. Reported melting temperatures present the maximum of the first derivative of the curve and are an average of the two measurements within a deviation of 1°C. T_m curves for each probe are shown in Supporting Information Figures S15–S20 in Ferreira et al. Reprinted from Ferreira et al.⁹⁹

3.6.4 Oncogene Information

Oncogene	Gene ID	HGNC	Ensembl	MIM	GRCh38.p13
BRAF	673	1097	ENSG00000157764	164757	GCF 000001405.39
KRAS	3845	6407	ENSG00000133703	19007	GCF 000001405.39
EGFR	1956	3236	ENSG00000146648	131550	GCF 000001405.39

Table S3.4: Summary of the oncogenes used for the FLEET assay and their human assemblies.

4 Mutations in microRNA-128-2-3p Identified with Amplification-Free Hybridization Assay

Sofie Slott,^a Cecilie Schiøth Krüger-Jensen,^a Izabela Ferreira da Silva,^{b,c,d} Nadia Bom Pedersen,^a and Kira Astakhova^{*a}

ABSTRACT. We describe a quantitative detection method for mutated microRNA in human plasma samples. Specific oligonucleotides designed from a Peyrard-Bishop model allowed accurate prediction of target:probe recognition affinity and specificity. Our amplification-free tandem bead-based hybridization assay had limit of detection of 2.2 pM. Thereby, the assay allowed identification of single-nucleotide polymorphism mismatch profiles in clinically relevant microRNA-128-2-3p, showing terminal mutations that correlate positively with inflammatory colitis and colorectal cancer.

4.1 Introduction

Single-nucleotide polymorphisms (SNPs) are fundamental drivers of adaptation and disease origins in human cells. Therefore, SNPs can be used for patient stratification, facilitating personalized medicine approaches¹³². Detection of SNPs in human RNA is more challenging compared to DNA, as the RNA sequence is relative short, has a higher chemical and enzymatic instability, and the RNA active 3D folding decreases binding of the detection reagents⁴⁹. In general, SNP detection reagents rely on matched probe:RNA oligonucleotide recognition, such as in sequencing and polymerase chain reaction (PCR). Sequencing is not suitable for large-scale screening, as next generation sequencing is limited by high cost and long assay time, while third generation sequencing does not meet specificity requirements⁴⁶. PCR restriction fragment length polymorphism (PCR-RFLP) has been used to identify SNPs in RNA. The method rely on PCR followed by enzymatic digestion to detect allelic variations by gel separation^{133–135}. The major limitation of this method is the need to identify and apply only SNPs that overlap restriction

enzyme recognition sites. As alternative approach, mismatch hybridization strategies for SNP detection has been developed in previous studies for longer RNA targets^{132,136}.

MicroRNA, also abbreviated miRNA, are typically 20-22 nucleotide long nucleic acids that regulate expression of mRNA. MiRNAs have a confirmed role in multiple cancer types including breast, colon, prostate and lung cancers¹³⁷⁻¹⁴¹, through regulation of for instance differentiation, apoptosis, migration and invasion¹³³. Verified by next-generation sequencing, mutations in miRNA correlate with their biogenesis and are linked to inflammatory and cancerous diseases²⁴. For example, SNPs rs11614913 and rs2910164 in miRNA-196 and miRNA-146a, respectively, are related to susceptibility to ulcerative colitis¹⁴². The inflammatory microenvironment influences miRNA expression, which in turn deregulate target gene expression and might cause cancer. As example, a recent study confirms that miRNA-26b impacts inflammation-associated colorectal carcinogenesis¹⁴³. As to the mechanism, SNPs might introduce structural changes into miRNAs impairing their function and thereby modulate gene expression levels, causing disease susceptibility or pathogenesis^{44,45,144}. In light of the growing interest for SNP profiling in miRNAs, accurate and sensitive detection methods are on high demand²⁴. Recently, Jia et al. applied a CRISPR/Cas12a RNA directed nuclease system, in combination with hybridization chain circuit, for reverse transcription-free and amplification-free miRNA detection⁶⁹. Although not sensitive to SNPs, the method could identify different miRNA in human cells in agreement with RT-qPCR, with a limit of detection down to 100 fM⁶⁹.

In previous work, we tested several miRNA in sera of human patients and in rodent model of colitis-associated colon cancer, using a tandem hybridization bead-based assay¹¹². We identified miR-128-2-3p as new potent biomarker for colon cancer progression. For the work, we used large DNA (>100kDa) as a scaffold for multiple fluorophore attachment, in order to boost the fluorescence read-out of the assay, and even with detectable photobleaching, reach the limit of target detection 10pM with a standard fluorometry¹¹¹. Confirming the clinical relevance of miRNA128-2, aberrant expression is associated with many types of human tumors¹⁴⁵. Among others, a prognostic marker potential of miRNA128-2 is identified for hepatocellular carcinoma¹⁴⁶. In addition, it induces chemoresistance in non-small cell lung cancer¹⁴⁷. A potential pitfall and a scientific challenge of any miRNA detection method is the design of the specific detection probes. DNA:RNA binding needs a high binding affinity and specificity prediction to make the assay specific and sensitive. Recently, we applied a unique Peyrard-Bishop mesoscopic model to design oncogene specific probes, which accounts for both hydrogen bonding and pi-stacking interactions between each individual nucleotide pair^{76,99,148}. The model included locked nucleic acid (LNA) enriched probes that were designed specifically to target the desired DNA and RNA. LNA is a nucleic acid analogue which improves the probes' affinity and specificity especially for DNA:RNA recognition¹¹². Using this model, we optimized oligonucleotide reagents so that the target SNP in coding RNA was discriminated with high specificity¹⁴⁸. In this work, we

hypothesized that an amplification-free assay with rationally designed LNA/DNA probes can be applied to rapid and accurate SNP profiling in human miRNA. We verified our hypothesis testing plasma sample using a bead-based fluorescence hybridization assay in a cohort of 24 colitis samples, 20 colorectal cancer (CRC) samples and 20 matched healthy controls.

4.2 Experimentals

4.2.1 Model

The Peyrard-Bishop model is described in details in SI. In brief, the model describes the DNA helix through a Hamiltonian, which contains a Morse potential describing the hydrogen bonds that connect each base-pair and a harmonic potential describing the stacking interaction of adjacent base-pairs^{76,126–128}. Details on temperature prediction, parameter and DNA:RNA binding optimization are also provided in SI. The experimental dataset of melting temperatures has 73 DNA:RNA hybrids sequences between 6-and 21-mer length at concentration of 100mM [Na+] and 8μM of strand.

4.2.2 Oligonucleotides

Biotinylated DNA and miR-128-2-3p were all purchased from IDT. The calf thymus DNA (CTD) binding linker was synthesized by solid-phase oligonucleotide synthesis. Details on linker synthesis and characterization is provided in SI.

4.2.3 UV melting studies

Capture probes and model miRNA targets (see sequences in Table S4.3) were annealed by mixing equal molar amounts in 1xPBS, pH 7.4, at 2μM final concentration, heating at 85°C for 10min, followed by cooling to room temperature over 1 h. UV melting experiments were carried out on JASCO V-730 instrument equipped with PAC-743 Peltier 8-position cell changer. Each melting was done in duplicate; T_m was determined as second derivative of absorbance at 260nm vs. temperature curve.

4.2.4 Fluorophore study in chip

For choosing the best dye for our miRNA detection assay, five different dyes were evaluated. EvaGreen (EG), AccuClear (AC), QuantiFluor (QF), Acridine Orange (AO) and Thiazole Orange (TO). Different dye concentrations were examined in order to investigate the intensity in presence of the CTD (booster). The samples were prepared by mixing 1xPBS with booster, and lastly the

dye was added and mixed by vortex. The booster used was calf thymus DNA (CTD; Sigma reagent no. D4522) with stock concentration at 4 mg/ml. The samples were covered in aluminum foil to prevent exposure to UV-light. 45µl of each sample was loaded in a one-chamber FLEET chip (design and production of the chips are described in SI) and analyzed by a Spectrofluorometer FS5 from Edinburgh Instruments with a SC-10 Front-Face Sample Holder module. Each test was run in duplicates. The CTD concentration in each sample was 2µM (10µL stock in 200µL sample) and dye concentrations varied as followed: EG, AC, QF: 0.5%, 1%, 2.0%, 2.5%, 3.0%, 4.0% and AO, TO: 1.0%, 2.5%, 3.0%, 4.0%, 5.0%, 7.5%, 10.0%. Blanks had no CTD with 4% or 10% dye.

In order to test the photostability of the dyes, the samples were remeasured seven times. Each dye was tested in duplicates. After remeasurement, the photobleaching was calculated in percentage by the formula: $\% \text{ photobleaching} = 100 - \left(\frac{\text{max last emission}}{\text{max first emission}} \right) \cdot 100$

To determine the limit of detection, we made a calibration curve for known amount of CTD with QuantiFluor fluorophore (CTD used in concentration range 0.1pM-1nM in presence of excess QuantiFluor fluorophore (100nM), in 1×PBS, pH 7.2). The fluorescence read out was obtained using Roche Light Cycler 480 equipment, at excitation/emission wavelengths 480-510nm/520-570nm.

4.2.5 Cohort information

De-identified human colitis plasma samples were obtained from Stanford University (USA) under Stanford Institutional Review Board (IRB protocol 28427). Human colon cancer (CRC) samples were provided by Danish Biobank Regionernes Bio- og GenomBank (RBGB) Herlev, under National Videnskabetisk Komité (National Research Ethics Committee) (NVK) number 21015520. All the samples were fully anonymized prior to analyses (Supplementary **Tables S4.8, S4.9**). EDTA tubes were used for human blood collection. Plasma was obtained by 15min centrifugation at 2,000 g in a bench top centrifuge (4°C) of the fresh whole blood. The obtained plasma samples were stored in aliquots at -80°C prior to usage.

4.2.6 miRNA analysis

4.2.6.1 Probe design

miRNA bead assay was conducted in duplicate following our published procedure¹¹¹. The design principle of the probes has been described earlier¹¹¹. miRNA target sequences were obtained from miR database (<http://www.mirbase.org/>). Complementary binary DNA probes were designed for each miRNA and enriched LNA to have high difference in melting temperature (T_m) for matched (mutated) vs mismatched (wildtype) target:probe duplex. Resulting probe sequences are conserved for human and mice samples and are given as: miRNA128-2-3p target code:

MIMAT0000424, wild type miRNA target: 5'-ucacagugaaccggucucuuu-3'. The linker was synthesized by solid-phase oligonucleotide synthesis, purified, and characterized as described in SI. The CTD-binding part of the detection probes has been obtained from the National Center for Biotechnology Information (NCBI) Nucleotide database (sequence ID GJ060426.1).

4.2.6.2 miRNA detection assay

For miRNA detection, streptavidin magnetic beads (NEB S1420S; 10 μ l, 4mg/mL) were placed in an Eppendorf tube and the clear supernatant was discarded using a magnetic separation rack. Then the beads were washed 3 times using each time 100 μ l 1 \times PBS, pH 7.2. Hereafter, the magnetic beads were resuspended in 1 \times PBS (20 μ l) and incubated with the miRNA-specific biotinylated capture probe (10 μ l, 1pmol) for 10 minutes on a shaker. When the incubation was finished the beads were washed with 1 \times PBS (2 times by 100 μ l) at RT and subsequently resuspended in 1 \times PBS (20 μ l). Next, plasma sample (10 μ l) or a known amount of synthetic miRNA target was added, and the resulting mixture was heated for 40 min at 65°C with repetitive vortexing every 10min, and subsequently cooled to 30°C over 30min followed by a wash with 1 \times PBS (2 times by 100 μ l) at 30°C. After cooling to RT, the CTD-signal boosting DNA (5pmol; 2 μ l) was added, followed by addition of the linker/detection probe (25 μ l; 2pmol), and 2 \times PBS (25 μ l) to the solution. The mixture was heated for 10min at 85°C and cooled to 16°C over 20min followed by a wash with 2 \times PBS (3 times with 100 μ l) at room temperature (23°C). The mixture was resuspended with 2 \times PBS (30 μ l), heated 10min at 92°C and the supernatant was extracted. The supernatant was placed in a solution with QuantiFluor (3 μ l, 20 \times stock) and 2 \times PBS (10 μ l). The resulting mixture was vortexed for 7min, followed by 5min at 92°C and finally cooled to room temperature (23°C) over 60min. Fluorescence read out was obtained using Roche Light Cycler 480 equipment, at excitation/emission wavelengths 480-510 nm/520-570 nm.

4.2.6.3 Quantification and Validation

To quantify the amount of miRNA captured on the mutation specific capture probe, we used a calibration curve (**Figure S4.6**) from CTD in QF on the fluorescence readout obtained from the Roche Light Cycler 480. This was converted to level of captured miRNA in each sample [nM] and are shown in **Table S4.10-S4.12** for CRC, colitis and healthy samples. The results of the bead-based assay for wild type miRNA were verified by commercial TaqMan RT-qPCR as described (ThermoFisher)⁴⁸. DNase-treated RNA extracted from plasma samples (Qiagen RNeasy kit, as suggested by manufacturer), was retrotranscribed with the SuperScriptTMIII First-Strand Synthesis System for RT-PCR (Invitrogen). RNA levels were analyzed by real-time quantitative TaqMan RT-qPCR (ThermoFisher), following the manufacturer's instructions. Reactions were performed with the LightCycler® 480 Instrument (Roche) in 384-multiwell plates. Primers for wildtype and mutated miR-128-2-3p (3' and 3'-1) were custom designed and provided in a mastermix by ThermoFisher. Thermal cycling was performed as follows: one pre-incubation cycle

at 95°C for 10 min (ramp rate: 4.8°C/s); 35 amplification cycles at 95°C for 10sec, 57°C for 30sec and 72°C for 5sec (ramp rates: 4.8, 2.2 and 4.8°C/s). The obtained Ct values for detected miRNA (wt, 3' and 3'-1 variants) were converted to concentrations by use of the standard curve (**Fig. S6.3.1**).

4.2.7 Statistical analysis

Statistical data analysis was performed by applying descriptive statistics (Students t test, Welch test and regression analyses, Pearson correlation, linear regression) in R. A P-value ≤ 0.05 was considered statistically significant.

4.3 Results

4.3.1 “Booster” dye optimization

We started with developing an ultra-sensitive amplification-free protocol that would allow detection of miRNA variants at low concentration. We choose the tandem hybridization bead based method, used in previous work, and optimized the protocol to get lower limit of detection (LOD)^{111,112}. Calf thymus DNA (CTD) was used as a scaffold for multiplying the number of fluorophores per target RNA:probe binding event¹¹¹. We selected a chip format for rapid screening of different dyes compatible with hybridization-based nucleic acid detection. The chip was prepared as a one-chamber device made of PMMA (described in SI). Using this convenient device, we rapidly investigated five fluorescent dyes that recognize DNA; EvaGreen, AccuClear, QuantiFluor, Acridine Orange and Thiazole Orange for photostability and emission intensity in the presence of CTD used as signal booster. **Figure 4.1A** shows representative emission spectrum of QuantiFluor. As shown in **Figure 4.1B**, QuantiFluor and Thiazole Orange emit the most florescence relative to their concentration. Photobleaching happens over multiple absorption- and emission-circles, where a fluorophore loses the ability to emit light¹⁴⁹. Photobleaching was examined by remeasuring the same samples seven times in duplicates. **Figure 4.1C** shows the percentage photobleaching observed a different dye concentration. AccuClear and QuantiFluor were found to be most photostable. Based on the data from the emission and photobleaching studies, we selected QuantiFluor as the “booster” dye for our miRNA detection assay, giving a limit of detected (LOD) of 2.2pM and bleaching 2.5% on average over seven exposures.

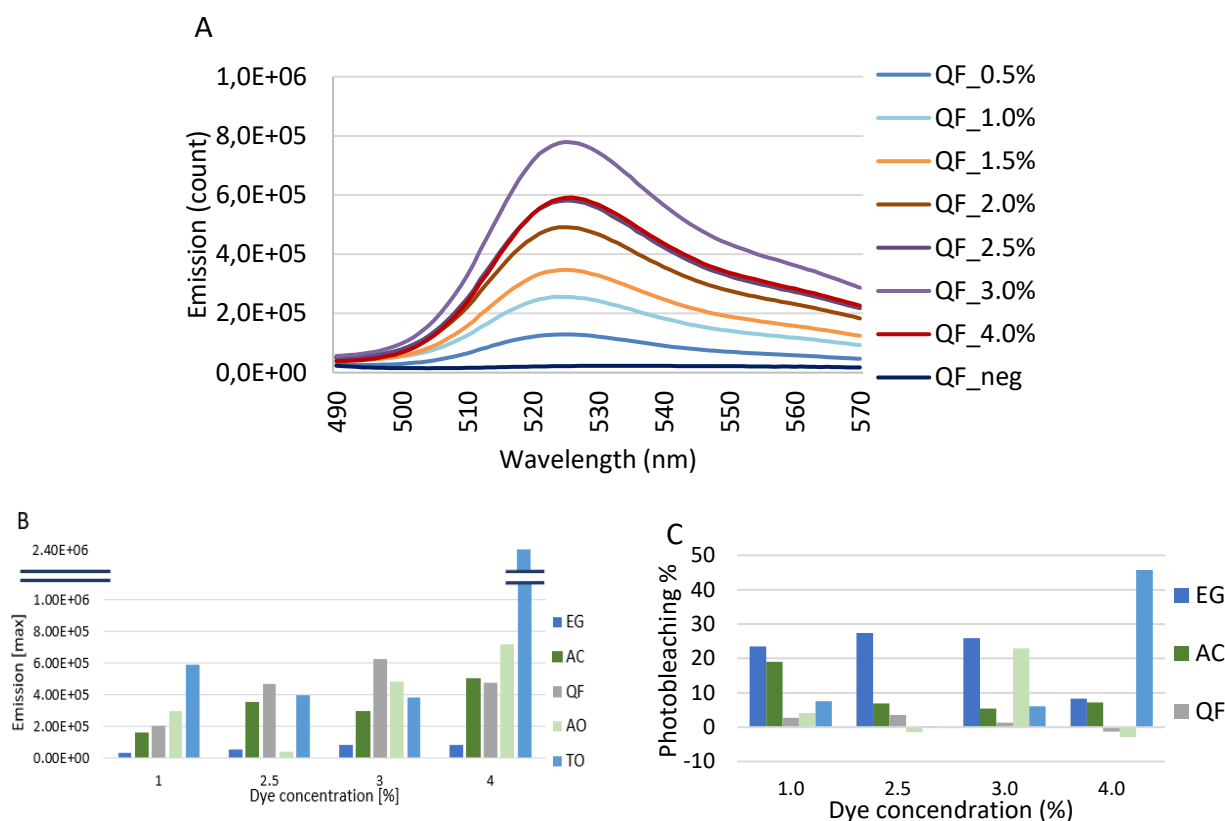


Figure 4.1: Fluorophore study in chip. Five dyes were evaluated, namely EvaGreen (EG), AccuClear (AC), QuantiFluor (QF), Acridine Orange (AO) and Thiazole Orange (TO). A) Representation emission spectrum of QF. B) Grouped plot of the maximum emission values for each dye in comparable concentrations. C) Dye photobleach (% from first to last measurement).

4.3.2 SNP probe design

After determining the optimal booster dye, we proceeded with analysis of miRNA in human plasma samples. For this, we selected miRNA128-2-3p as a potent biomarker with upregulation in colitis and colorectal cancer¹¹². The probes for the assay were designed as a tandem hybridization system, where biotinylated capture probe attached to the magnetic bead surface was a mutation-specific LNA-enriched oligonucleotide (**Figure 4.2**). Secondary probe, called linker, binds to the 5' end of the miRNA and contains sequence that specifically interacts with the signal booster. It has sequence: 5'-T+C+A+CT+GT+GATTTTGAT+GGG+AATAC+CAGACC-3'. The first part (dashed) adjusts to 5'-end of the target miRNA with a predicted T_m of 45°C°. The underlined 3'-end binds to CTD. LNA nucleotides are indicated with a plus (+) in front of the corresponding nucleotide letter. As a proof of principle, we designed 12 mutation specific capture probes of 12 nucleotides in length, recognizing mutations at 3' terminus and 3 adjacent positions (3', 3'-1, 3'-2,

3'-3) of the target miRNA. The probe sequences were designed using a parameterization of the Peyrard-Bishop model for LNA-modified probes. The model describes the sequences through the hydrogen-bonds between each base-pair and their stacking interaction^{76,148}. The design allowed up to four LNA additions per sequence to reach improved binding affinity to target miRNA. We aimed at a drop in the melting temperature in the presence of mismatch in the target. Final probes and their predicted discrimination of mismatch (ΔT_m) are shown in **Table 4.1**. The design was successful in terms of discriminating SNP; with melting temperature difference (ΔT_m) for match vs. mismatch in range 8.8°C – 17.2°C (**Table 4.1**).

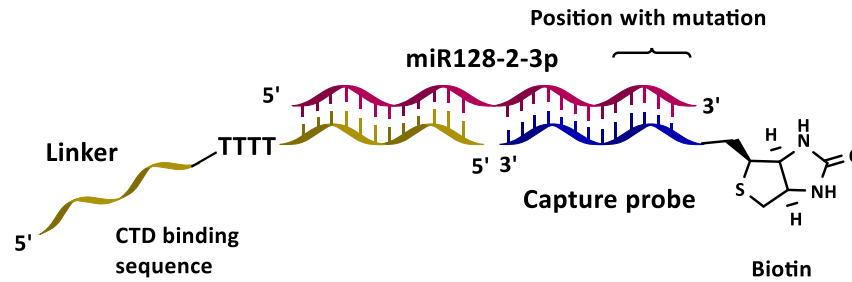


Figure 4.2: Design of tandem hybridization probes for this study. Biotinylated mutation specific capture probe (blue), miRNA128-2-3p or variant (pink), Linker (yellow) with booster binding part. CTD = calf thymus DNA (booster).

Table 4.1. Mutation specific capture DNA/LNA probes targeting miRNA128-2-3p applied in this study; predicted and experimental difference in melting temperatures (ΔT_m) between fully matched (mutant) and mismatched (wild type) RNA:capture probe complexes.

#	Sequence, 5' → 3'	Position of mutations	ΔT_m , pred, °C	ΔT_m , exp, °C
C1	Bt-Sp9-d(AAAGAG+AC+CG+GT)	None, wt	13.3	17.3
C2	Bt-Sp9-d(<u>T</u> AAGAG+AC+CG+GT)	3', U>A	11.3	16.4
C3	Bt-Sp9-d(<u>C</u> AAGAG+AC+CG+GT)	3', U>G	11.3	12.8
C4	Bt-Sp9-d(<u>G</u> AAGAG+AC+CG+GT)	3', U>C	11.3	6.6
C5	Bt-Sp9-d(A <u>T</u> AGA+GACC+G+GT)	n-1, U>A	10.0	18.4
C6	Bt-Sp9-d(A <u>C</u> AGA+GACC+G+GT)	n-1, U>G	10.0	10.2
C7	Bt-Sp9-d(+A <u>G</u> A+GAGA+CCGGT)	n-1, U>C	10.3	25.5
C8	Bt-Sp9-d(AA <u>T</u> +GAG+A+CCGGT)	n-2, U>A	10.8	NM
C9	Bt-Sp9-d(AA <u>C</u> +GAGA+CCGGT)	n-2, U>G	10.8	NM
C10	Bt-Sp9-d(AAG+GAGA+CCGGT)	n-2, U>C	10.8	NM
C11	Bt-Sp9-d(AAA+ <u>T</u> +AGACC+GGT)	n-3, C>A	8.8	NM
C12	Bt-Sp9-d(AAA <u>C</u> AGA+CCG+GT)	n-3, C>G	17.2	NM
C13	Bt-Sp9-d(AAA <u>A</u> AG+A+CCG+GT)	n-3, C>U	14.6	NM
C14	Bt-Sp9-d(AGCGCGGATAAA)	scramble	14.6	NM

*Wt = wild type. Bt = biotin; Sp9 = spacer C9H18; + =LNA; pred=predicted, exp=experimental, NM = not measured. Position opposite to the mutated nucleotide in the target miRNA is underlined. Wild-type miRNA128-2-3p: 5'-r(UCACAGUGAACCGGUCUCUUU)-3'.

4.3.3 Tandem hybridization bead-based assay – Detecting SNPs in human plasma miR-128-2-3p

Prior to the assay, melting temperatures of capture probes were evaluated by a thermal denaturation study. We analyzed C1-C7 with complement model miRNA targets (M1-M7, **Table S4.4**) and results are seen in **Table S4.5**. Difference in T_m between two independent measurements did not exceed 1.5°C. The experimental ΔT_m of matched vs. mutant was found to range between 6.6°C – 25.5°C (**Table 4.1**, ΔT_m exp.), which is a slightly broader range than predicted. However, these UV melting temperature studies confirmed that the designed capture probes could effectively discriminate between the wild type miRNA128-2-3p and its mutated analogues. For the amplification-free assay, biotinylated capture probes C1-C14 were attached to streptavidin-coated magnetic beads in individual reaction tubes (**Figure 4.3**). We used elevated temperature (30°C) in the following washing step to ensure that unspecific capture probe:target binding did not take place. After sequential incubation with patient plasma sample, linker probe and booster were added. Followed by several washes, the complex was denatured and the supernatant transferred to a QuantiFluor solution after which the fluorescence signal was detected (**Figure 4.3**).

A fluorescence calibration curve was established to convert optical read out of the amplification-free assay into absolute amounts of either wild type or mutated miRNA. The absolute amounts of wild type and mutated miRNA obtained for human plasma samples are shown in **Figure 4.4**. For the wild type target, the assay was validated using RT-qPCR as we reported before¹¹¹. Regression analysis gave the following model of the concentration of miRNA128-2-3p in nM for all tested samples: $c(bead) = 0.49 \cdot c(qPCR)$, $P\text{-val} = 3.2 \cdot 10^{-9}$, $r^2 = 0.43$, where bead refers to our bead-based hybridization assay. The model reveals a higher (p-value= $7.7 \cdot 10^{-4}$) wild type miRNA titer determined by RT-qPCR.

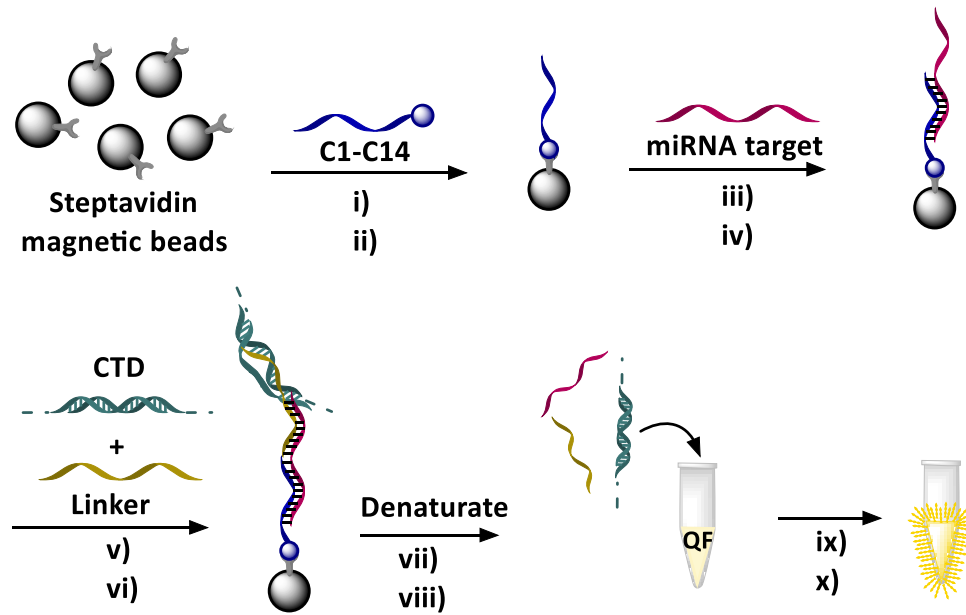


Figure 4.3: Procedure of miRNA detection bead-based assay: i) Streptavidin magnetic beads were incubated with biotinylated mutation specific capture probes (C1-C14) followed by ii) washing, iii) incubation with samples containing miRNA128-2-3p and variants then iv) washing. v) Beads were then incubated with CTD, linker and buffer followed by vi) washing and vii) denaturation by heating to 92°C in 10min, and viii) the supernatant transferred to QuantiFluor (QF) solution, ix) that was incubated at 92°C in 5min and cooled to 23°C over 60min. Finally, x) the concentration of miRNA128-2-3p and variants were quantified by fluorescence detection.

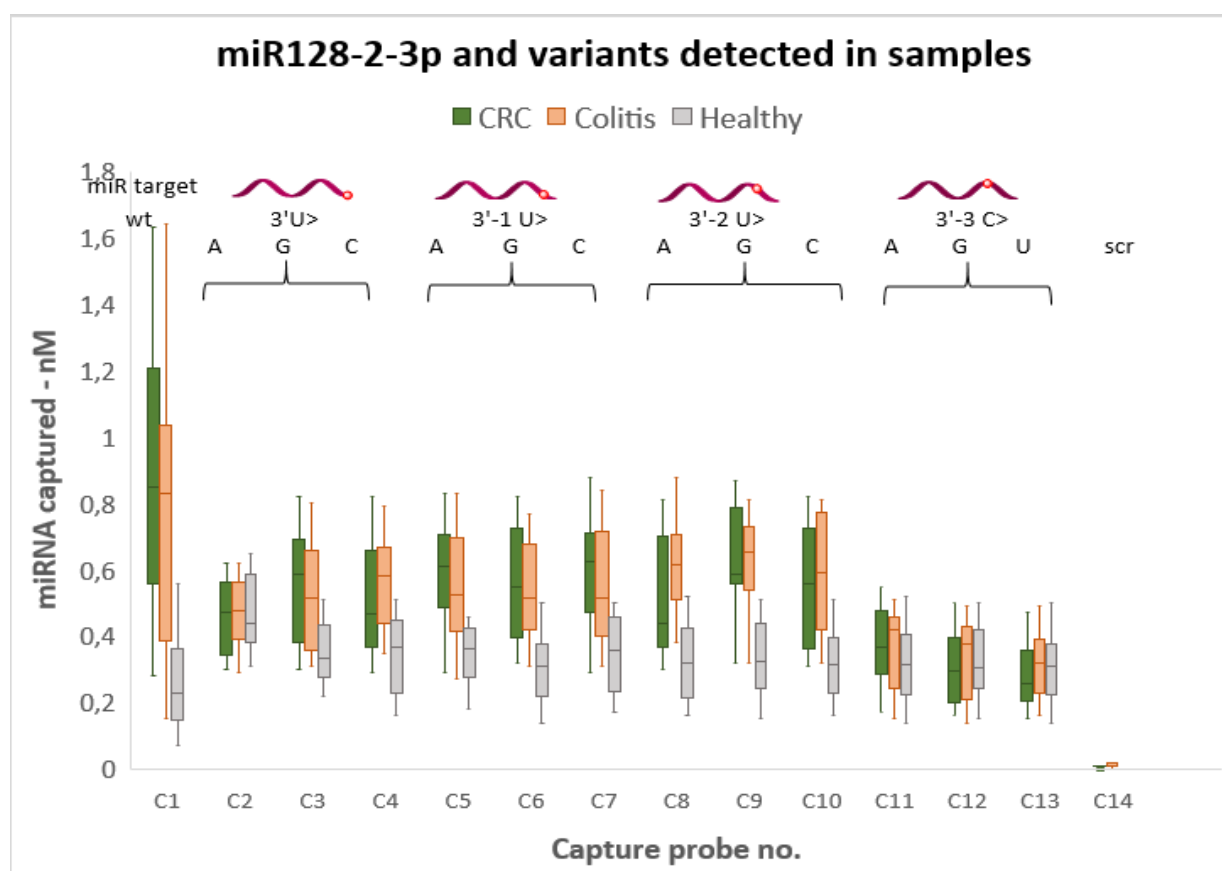


Figure 4.4: Results of amplification-free hybridization assay in human colorectal cancer (green), colitis (orange) and healthy (grey) plasma samples – Box-whisker plot with outliers. miRNA targets are illustrated with position of mutation (red dot). Letters show the individual modifications, correlating with the capture probe numbers (C2-C13); wt = wild type (C1); scr = scrambled capture probe (C14).

According to one-way ANOVA, there is a statistically significant difference among all groups in diseased samples (CRC, Colitis), and not in healthy controls (**Figure 4.4**). Notably, levels of mutated miRNA are higher in the diseased samples than healthy controls. It indicates an association between SNP mutations in miR-128-2-3p and diseases colitis and colorectal cancer. Next, there is a high similarity in miRNA mutation pattern between colitis and colorectal cancer, with slightly higher levels of mutated miRNA in the former. In both diseased groups, levels of terminal mutated miR-128-2-3p in the first three positions (C2-C10) are higher than for healthy controls. In particular of 3'-2 U>G and 3'-2 U>C mutations gives the highest levels. Opposite, for 3'-3 position the levels of mutated miRNA-128-2-3p are lowest. Using a QuantiFluor calibration curve, LOD of the assay is determined to be 2.2pM RNA target which is 5 times lower compared

to previously used EvaGreen^{111,112}. High specificity of the assay is confirmed with no binding for scrambled control probe C14.

4.3.4 Correlation with other clinical parameters

To further explore the biological role of the SNP mutations in miRNA128-2-3p, we performed Pearson correlations of clinical features for subjects with obtained levels of miRNA (**Table S4.18**). Relatively weak correlations were observed, with Pearson coefficients in the range -0.34 to +0.38. Interestingly, multiple correlations were observed for same few miRNA variants, namely C4, C6, C7 and C10. Age correlated negatively with C4 and C12 levels (Pearson coefficient -0.21 and -0.3, respectively); as to gender, there were no correlations. Disease type and stage correlated negatively with levels of several mutated miRNAs, also with wild type (C1, C4 and C10). A positive correlation was seen for history of ulcerative colitis with C8 (Pearson coefficient 0.38); and for hemoglobin levels correlating positively with C4 and C10 (Pearson coefficient 0.29 and 0.27, respectively), and erythrocyte sedimentation rate (ESR), correlating positively with C5 (Pearson coefficient 0.21). Last, we observed positive correlation of mutated miRNA-128-2-3p (C6, C7 and C13), with race of subjects, no-White group having higher correlation with the mutated miRNA.

4.4 Discussion

In this study, we optimize a bead-based hybridization assay to accurately detect SNPs in human miRNA128-2-3p. From examinations of five different dyes, QuantiFluor was chosen for the miRNA detection bead-based assay as it had the best emission qualities and photostability. By this, we achieved an optimized LOD from 10pM to 2.2pM¹¹¹. In the literature, the majority of existing miRNA detection methods require enzymes for signal amplification such as RTPCR, rolling circle amplification, strand displacement amplification and exonuclease/nicking endonuclease-assistant amplification⁶⁶. Only few amplification-free methods sensitive to miRNA SNPs have been developed, and even less has been applied to detect SNPs in miRNA^{66,67,150}. In 2014, Qiu et al. described the use of a hybridization chain reaction to fluorescently identify let-7a on SNP level using two different silver nanocluster hairpin oligomers⁶⁶. Caputo et al. designed a miRNA detection platform using microgels and molecular beacon stem loops, that when conjugated together, could catch miRNA in samples⁶⁷. The biosensor showed SNP selectivity and low sensitivity (nanomolar to picomolar order, LOD 10fM, detection time 1h). The miRNA detection methods described in Caputo et al. and Qiu et al. both show SNP selectivity. However, the methods is not applied for detection of specific miRNAs SNPs in the biological samples and are only analyzing the wild-type miRNA^{66,67}. A recent study by Xia et al. applied the SYBR green

fluorophore for SNP detection. In the assay, graphene oxide was added to prevent SYBR green dsDNA insertion, quenching the fluorescent signal and thus allowing SNP detection¹⁵¹. The method has LOD of 1nM, but was used for DNA rather than (mi)RNA targets, which were synthetic rather than biological¹⁵¹. It is well known that mutations appear at an extremely low frequency in genome and transcriptome^{152,153}. Although an amplification-free approach can efficiently reduce the complexity of an experiment system, the detection sensitivity of these assays can in general not meet the requirements of biological analysis and clinical applications. In this assay, the detection limit is determined to be 2.2pM, which is poorer than isothermal exponential amplification reaction (EXPAR) with chimeric DNA probe-aided precise RNA disconnection (1fM)¹⁵⁴ and tandem gene amplification (8.3fg)¹⁵⁵. Nevertheless, given relatively high miRNA concentration in the samples we tested (0.07-1.64nM), applying our method is feasible (Supplementary **Tables S4.10-S4.12**).

Using our probes in an amplification-free assay, we show that wild type miRNA128-2-3p is present in a significantly higher amount in colitis and CRC patients than in healthy controls and therefore, the miRNA can be used as a biomarker for colon cancer progression. Furthermore, we were able to quantify the amount of SNP variants in several positions of miRNA128-2-3p obtained from the samples from colitis and CRC patients. To the best of our knowledge, SNP in miRNA128-2-3p associated with colitis or CRC have not yet been reported. A risk of false positive signal is considerable when detecting miRNA SNP such as with miR-141 and miR-200a reported by Metcalf et al.¹⁵⁰. However, our data can be fitted with TaqMan qPCR data, confirming accuracy of the developed assay. For the RT-qPCR, a higher amount of wild-type miRNA128-2-3p was detected compared to our bead-based assay, which lowers the risk of false positive samples. However, it also indicates that the calibration and accuracy of the assay can be optimized further.

Applying our bead-based assay in human plasma samples, we detect a significant difference in mutated miRNA levels of diseased samples, but not in healthy controls. It indicates that healthy individuals have baseline levels of SNP miRNA128-2-3p, some of which gets elevated upon disease state. Therefore, the miRNA128-2-3p SNPs giving the best discrimination of healthy and disease individuals can be selected as biomarkers of CRC and colitis. Our study indicates that SNP mutations 3'-2U>G and 3'-2U>C (capture probes C9 and C10) are promising biomarkers of CRC (median difference is 44% for both, P-values $3.1 \cdot 10^8$ and $1.2 \cdot 10^5$, respectively). Likewise, SNP mutations 3'-2U>A and 3'-2U>G (capture probes C8 and C9) discriminates colitis well which median differences of 48% (p-value: $1.6 \cdot 10^{-9}$) and 50% (p-value: $1.6 \cdot 10^{-8}$), respectively. In addition, Person correlation was positive between 3'-2U>A (C8) and history of ulcerative colitis (Pearson coefficient: 0.38), while 3'-2U>C (C10) positively correlated with hemoglobin (Pearson coefficient: 0.27), while negatively with disease type and stage (Pearson coefficient: -0.23). Importantly, the 3'-2 mutations were not correlated with either age or gender. Taken together, 3'-2 SNP mutations

of miRNA128-2-3p are promising as biomarkers of CRC and colitis. However, the finding should be verified in a larger cohort.

This work demonstrates the utility of a tandem hybridization bead-based assay to detect SNP mutation in specific miRNA miRNA128-2-3p. However, we predict a potential to expand the technology to detection of various clinical and biological relevant RNA biomarkers. The developed Peyrard-Bishop model offers a tool for design of high-sensitivity probes, which can adapt the assay to screening of any miRNA SNP, potentially allowing early disease diagnosis.

4.5 Conclusions

In summary, we designed specific hybridization probes to detect SNPs in miRNA128-2-3p. We applied the Peyrard-Bishop mesoscopic model to design LNA enriched capture probes with high specificity for miRNA128-2-3p and its mutants. We showed high discrimination between match and mismatch probes with a ΔT_m range between 6.6-25.5°C. The mutation-specific capture probes were incorporated into bead-based amplification-free detection platform. Using the assay, we successfully quantified levels of miRNA128-2-3p and its mutated variants in plasma samples from colitis and colorectal cancer patients. We improved the LOD from 10pM to 2.2pM using the fluorescent dye QuantiFluor. Using the bead-assay we demonstrated a significantly higher level of the wild type miRNA128-2-3p in colitis and colorectal cancer than in healthy. Moreover, we observed a generally higher amount of mutated miRNA128-2-3p in colitis and CRC than in healthy controls, except for mutations in the 3'-3 position. Our approach can be used to detect various biologically and clinically relevant RNA biomarkers, which might be useful in fast and efficient early stage diagnosing or screening.

Author Information

- [a] Department of Chemistry, Technical University of Denmark, 206-207 Kemitorvet, 2800 Kgs. Lyngby, Denmark. *E-mail: kirras@kemi.dtu.dk
- [b] Programa Interunidades de Pós-Graduação em Bioinformática, Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brazil.
- [c] Departamento de Física, Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brazil.
- [d] Bioinformatics Core, Luxembourg Centre For Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, House of Biomedicine II, 6, avenue du Swing, L-4367 Belvaux.

Author Contribution

All authors participated in assay design and data handling; SS, CSKJ and KA conducted the assay and experimental procedure; IF conducted the thermodynamic calculations and predictions on probe design; KA conducted statistical analyses; SS, NBP and KA drafted the paper; all authors participated in revising the paper prior to submission.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by DTU Enable Foundation. We thank Stanford University, School of Medicine, CA, USA, and Danish Biobank Regionernes Bio-og GenomBank (RBGB) Herlev, for providing us with colitis, colon cancer plasma samples and healthy controls.

4.6 Supporting Information

4.6.1 Mesoscopic Modelling

4.6.1.1 Model

The Peyrard-Bishop model describes the DNA helix through a Hamiltonian, which contains a Morse potential describing the hydrogen bonds that connect each base-pair and a harmonic potential describing the stacking interaction of adjacent base-pairs^{76,126–128}.

$$U_{i,i+1} = \frac{k_{\alpha,\beta}}{2} (y_i - y_{i-1})^2 + D_{\alpha} (e^{-y/\lambda_{\alpha}} - 1)^2, \quad (1.1)$$

The equation above describes the interaction of the i th base pair with its nearest-neighbour, $i + 1$. Where, D_{α} and λ_{α} describes the depth and width of the i th base-pair of type α , respectively. An elastic constant $k_{\alpha,\beta}$, is used to describe the coupling between nearest neighbours base-pairs and the coordinate y represents the relative displacements between the bases.

The sum for the Eq. (1.1) over all possible N base-pairs is carried out using its partition function:

$$Z_y = \int_{y_{min}}^{y_{max}} dy_1 \int_{y_{min}}^{y_{max}} dy_2 \cdots \int_{y_{min}}^{y_{max}} dy_N \times \prod_{n=1}^N e^{-\beta U(y_i, y_{i+1})} \quad (1.2)$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant and T the absolute temperature. Subsequently, the integral over all possible configurations of base pair displacements, y_i is performed. From the partition function, Eq. (1.2) an adimensional index τ is derived and it is directly correlated with the experimental melting temperatures. Furthermore, the average displacement, $\langle y_m \rangle$, at the m th position in the sequence can be derived from:

$$\langle y_m \rangle = \frac{1}{Z_y} \int_{y_{min}}^{y_{max}} dy_1 \int_{y_{min}}^{y_{max}} dy_2 \cdots \int_{y_{min}}^{y_{max}} dy_N \times y_m \prod_{n=1}^N e^{-\beta U(y_i, y_{i+1})} \quad (1.3)$$

4.6.1.2 Temperature Prediction

Considering a set $P = \{p_1, p_2, \dots, p_F\}$ of Morse potentials, D , and stacking parameters, k , an index $\tau_i(P)$ is calculated for each sequence i using the partition function, Eq. (1.2), for the Peyrard-Bishop Hamiltonian. Therefore, the melting temperature, $T'_i(P)$ for each parameter set, P , is derived from the equation:

$$T'_i(P) = a_0(N) + a_1(N)\tau_i(P) \quad (1.4)$$

where N is the length of the sequence.

4.6.1.3 Parameter Optimization

The optimization process consists in minimizing the merit function

$$\chi_j^2 = \sum_{i=1}^N [T'_i(P_j) - T_i]^2 \quad (1.5)$$

where $T'_i(P_j)$ is the calculated melting temperature from Eq. (1.4), T_i is the experimental melting temperature and N is the number of melting temperatures in the dataset. The search for the minimum is made through a downhill simplex method¹⁵⁶, and the procedure is made j times. In each j , a new set of Morse potentials, D , and stacking parameters, k is used. The problem of local minima is avoided repeating the minimization several times, starting over with different initial parameters following the procedures described in ^{127,157}. The final parameters shown here are those for the lowest overall χ^2 , and we estimate the parameter uncertainty from the standard deviations over the selected minimizations.

Additionally, we also use as a quality parameter the average prediction difference

$$\langle \Delta T \rangle = \frac{1}{N} \sum_{i=1}^N |T'_i - T_i|. \quad (1.6)$$

4.6.1.4 DNA:RNA Optimization

The partition function was integrated up to 400 points in an interval between $y_{min} = -0.1$ nm and $y_{max} = 20$ nm with an eigenvalue cut-off of 10 (Eq. (22) of Kawasaki et al.¹⁵⁸).

Seed parameters

In all optimizations the initial parameter p_i is varied randomly in an interval

$$p_i \in [(1 - f)s_i, (1 + f)s_i] \quad (1.7)$$

that is, within a fraction $\pm f$ of a seed value s_i .

For instance, $f = 0.1$, which results in the interval $[0.9s_i, 1.1s_i]$. As initial parameters for Morse potentials, D , and stacking parameters, k , were used generic parameters¹⁵⁷. The minimized Morse potentials and stacking parameters are displayed in the **Table S4.1**.

k (eV/nm ²)		D (meV)
dArU-dArU: 0.75(0.59)	dArU-dCrG: 7.09(1.94)	dArU: 20.75(12.42)
dArU-dGrC: 2.69(0.75)	dArU-dTrA: 4.42(1.63)	dCrG: 64.23(12.46)
dCrG-dArU: 3.77(1.30)	dCrG-dCrG: 2.48(0.63)	dGrC: 68.87(14.07)
dCrG-dGrC: 1.51(0.51)	dCrG-dTrA: 2.67(0.66)	dTrA: 40.69(12.51)
dGrC-dArU: 2.24(0.56)	dGrC-dCrG: 2.07(0.69)	
dGrC-dGrC: 1.47(0.73)	dGrC-dTrA: 1.93(0.72)	
dTrA-dArU: 1.34(0.90)	dTrA-dCrG: 2.98(0.55)	
dTrA-dGrC: 2.34(0.88)	dTrA-dTrA: 2.47(0.76)	

Table S4.1: DNA:RNA stacking (k) and Morse (D) parameters at 100mM [Na⁺] estimated uncertainty is shown in compact notation.**Experimental Data**

The dataset of melting temperatures has 73 DNA:RNA hybrids sequences between 6-and 21-mer length at concentration of 100mM [Na⁺] and 8 μ M of strand. The **Table S4.2** shows all the sequences with the experimental and predicted melting temperatures.

5' \rightarrow 3'	3' \rightarrow 5'	T_{exp}	T_{pre}
r(GGUCCG)	d(CCAGCG) ^e	27.2	26.52
r(CGGACC)	d(GCCTGG) ^e	26.1	25.41
r(GCCGUGAG)	d(CGGCACTC) ^e	41.2	40.92
r(GAGCCGUG)	d(CTCGGCAC) ^e	41.5	40.92
r(GUCAGACU)	d(CAGTCTGA) ^e	29.7	29.17
r(GACAGUCU)	d(CTGTCAGA) ^e	30.1	29.17
r(GAACUGCC)	d(CTTGACGG) ^e	33.5	31.96
r(GGCAGUUC)	d(CCGTCAAG) ^e	33.8	31.86
r(GCACAGCC)	d(CGTGTCGG) ^a	35.6	39.90
d(GCACAGCC)	r(CGUGUCGG) ^a	37.2	40.69
r(CCUUCCCUU)	d(GGAAGGGAA) ^a	20.5	23.42
r(UUCCCUUCC)	d(AAGGGAAGG) ^a	14.9	22.74
d(CCTTCCCTT)	r(GGAAGGGAA) ^a	44.8	41.57
d(TTCCCTTCC)	r(AAGGGAAGG) ^a	44.2	42.03
r(GCCAGUUA)	d(CGGTCAATT) ^d	30.6	29.92
r(GCGAUCGGA)	d(CGCTAGCCT) ^e	43.5	41.82
r(GCCAGUAGG)	d(CGGTCATCC) ^e	42.6	42.87
r(GCUCUCUGGC)	d(CGAGAGACCG) ^a	40.8	42.33
d(GCTCTCTGGC)	r(CGAGAGACCG) ^a	50.9	51.22
r(GAAGAGAAGC)	d(CTTCTCTTCG) ^b	46.9	43.10
d(GAAGAGAAGC)	r(CUUCUCUUCG) ^b	23.7	22.87
r(GUUCAAUACG)	d(CAAGTTATGC) ^e	27.5	28.24
r(AGGAUGACCG)	d(TCCTACTGGC) ^e	45.9	45.35
r(CGCUUGUUAC)	d(GCGAACAATG) ^e	33.1	29.68
r(GUAACAAGCG)	d(GTGAACAATG) ^e	39.2	39.10
r(CACUUGUUAC)	d(GTGAACAATG) ^e	28.1	25.17
r(AAUCUGGCCA)	d(TTAGACCGGT) ^e	42.8	40.06
r(AUGGCUCCAA)	d(TACCGAGGTT) ^e	40.1	40.06
r(GGGGAACAAGG)	d(CCCCTTGTTC) ^e	54.3	52.93
r(UUCACCUGGUC)	d(AAGTGGACCAG) ^e	45.3	40.92
d(TCCCTCCTCTCC)	r(AGGGAGGAGAGG) ^a	61.4	60.89

r(GGCAGGAAUCCG)	d(CCGTCCTTAGGC) ^e	56.8	54.94
r(GGAAUCAGGCCG)	d(CCTTAGTCCGGC) ^e	56.3	54.94
r(UAUCUUCCGAAU)	d(ATAGAAGGCTTA) ^e	30.2	32.46
r(UAUCCUUCGAAU)	d(ATAGGAAGCTTA) ^e	29.6	32.47
r(AAUGGAUUACAA)	d(TTACCTAATGTT) ^e	36.3	34.51
r(AUUGGAUACAAA)	d(TAACCTATGTTT) ^e	36.2	34.50
r(CCUGGAAUCCAA)	d(GGACCTTAGGTT) ^e	48.2	46.79
r(GGCUCAAUUGAC)	d(CCGAGTTAACTG) ^e	45.2	43.38
r(CGGCCUUGAUCC)	d(GCCGGAAGTACG) ^e	51.9	49.57
r(CGGAUUCUGCC)	d(GCCTAAGGACGG) ^e	50.3	49.58
r(UCCGAAUUAUCU)	d(AGGCTTAATAGA) ^e	35.8	32.47
r(AGAUAUUCGGA)	d(TCTATTAAGCCT) ^e	35.5	37.62
r(GCUUCUCUCUUC)	d(CGAAGAGAGAAG) ^e	31.5	31.17
r(GAAGAGAGAAGC)	d(CTTCTCTCTTCG) ^e	54.0	51.39
r(UCGUUCUUGUCU)	d(AGCAAGAACAGA) ^e	36.4	35.71
r(AGACAAGAACGA)	d(TCTGTTCTTGCT) ^e	47.6	47.45
r(AUUGGAUACAAA)	d(TAACCTATGTTT) ^d	35.5	34.50
d(GAGCTCCCAGGC)	r(CUCGAGGGUCCG) ^a	60.3	57.00
r(GAGCUCCCAGGC)	d(CTCGAGGGTCCG) ^a	56.7	57.95
r(UCCCUCCUCUCC)	d(AGGGAGGAGAGG) ^a	43.4	45.41
r(GUUAGCGUACGC)	d(CAATCGCAATGCG) ^e	45.0	46.20
r(GCGUUUACGUAGC)	d(CGCAAATGCATCG) ^e	47.8	46.19
r(UCACGUAGUCGUAU)	d(AGTGCATCAGCATA) ^e	49.8	49.57
r(UGUACGUCACAACUA)	d(ACATGCAGTGTTGAT) ^a	49.2	49.24
r(UAUACAAGUUAUCUA)	d(ATATGTTCAATAGAT) ^a	35.9	34.05
r(CGACUAUGCAAAAAC)	d(GCTGATACGTTTTTG) ^a	47.3	48.47
d(TAGTTATCTCTATCT)	r(AUCAUAGAGAUAGA) ^a	45.4	43.28
d(TGTACGTCACAATA)	r(ACAUGCAGUGUUGAU) ^a	50.6	50.64
d(TATACAAGTTATCTA)	r(AUAUGUCAAUAGAU) ^a	35.2	35.29
d(CGACTATGCAAAAAC)	r(GCUGAUACGUUUUUG) ^a	39.0	41.13
r(UAGUUAUCUCUAUCU)	d(ATCAATAGAGATAGA) ^a	34.9	34.45
d(CGACTATGCAAGTAC)	r(GCUGAUACGUUCAUG) ^c	45.1	48.84
d(CGCAAAAAAAAAAACGC)	r(GCGUUUUUUUUUUGCG) ^a	28.7	30.70
r(CGCAAAAAAAAAAACGC)	d(GCGTTTTTTTTTGCG) ^a	50.2	51.44
d(GGACCGGAAGGTACGAG)	r(CCUGGCCUCCAUGCUC) ^c	57.0	57.99
d(CTCGTACCTTCCGGTCC)	r(GAGCAUGGAAGGCCAGG) ^a	64.8	65.29
r(CUCGUACCUUCCGGUCC)	d(GAGCATGGAAGGCCAGG) ^a	56.0	57.99
r(CUCGUACCUUCCGGUCC)	d(GAGCATGGAAGGCCAGG) ^a	56.8	55.89
d(CTCGTACCTTCCGGTCC)	r(GAGCAUGGAAGGCCAGG) ^a	65.2	64.61

d(CTCGTACCATTCCGGTCC)	r(GAGCAUGGUAAGGCCAGG) ^c	63.7	64.22
r(GCCGAGGUCCAUGUCGUACGC)	d(CGGCTCCAGGTACAGCATGCG) ^a	68.1	66.04
d(GCCGAGGTCCATGTCGTACGC)	r(CGGCUCCAGGUACAGCAUGCG) ^a	68.2	64.98

Table S4.2: Sequences of DNA:RNA hybrids and their measured (T_{exp}) and predicted melting temperatures (T_{pre}). Main strands are shown on top of each row in 5'→3' direction, while complementary strands are in the bottom of the same row in 3'→5' direction. The d and r before the oligonucleotide mean DNA and RNA sequence, respectively. The final quality parameters of the optimization were $\langle \Delta T \rangle = 1.65$ °C and $\chi^2 = 330.58$ °C². ^a Sequences collected from ¹⁵⁹. ^b Sequences collected from ¹⁶⁰. ^c Sequences collected from ¹⁶¹. ^d Sequences collected from ¹⁵⁸. ^e Sequences collected from ¹⁶².

4.6.2 Oligonucleotides

4.6.2.1 Purchased Oligonucleotides

Biotinylated DNA and miR-128-2-3p were all purchased from IDT. Sequences are shown in Table S2.1. miR128-2-3p target and capturing probes (C1-C14) were used for miRNA SNP bead-assay detection. Model miR128 targets (M1-M7) and capturing probes (C1-C7) were used for UV melting studies (T_m). The CTD binding linker was synthesized by solid-phase oligonucleotide synthesis.

Name	5' → 3'	Name	5' → 3'
miR-128-2-3p	r(UCACAGUGAACCGGUCUCUUU)	C11	Bt-Sp9-d(AAA+T+AGACC+GGT)
C1 (wt)	Bt-Sp9-d(AAAGAG+AC+CG+GT)	C12	Bt-Sp9-d(AAA <u>C</u> AGA+CCG+GT)
C2	Bt-Sp9-d(TAAGAG+AC+CG+GT)	C13	Bt-Sp9-d(AAAA <u>A</u> AG+A+CCG+GT)
C3	Bt-Sp9-d(<u>C</u> AAGAG+AC+CG+GT)	C14 (scr)	Bt-Sp9-d(AGCGCGGATAAA)
C4	Bt-Sp9-d(<u>G</u> AAGAG+AC+CG+GT)	miR-128 M1	r(ACCGGUCUCUUU)
C5	Bt-Sp9-d(AT <u>A</u> GA+GACC+G+GT)	miR-128 M2	r(ACCGGUCUCUU <u>A</u>)
C6	Bt-Sp9-d(A <u>C</u> AGA+GACC+G+GT)	miR-128 M3	r(ACCGGUCUCUU <u>G</u>)
C7	Bt-Sp9-d((+A <u>G</u> A+GAGA+CCGGT)	miR-128 M4	r(ACCGGUCUCUU <u>C</u>)
C8	Bt-Sp9-d(AA <u>T</u> +GAG+A+CCGGT)	miR-128 M5	r(ACCGGUCUCU <u>A</u> U)
C9	Bt-Sp9-d(AA <u>C</u> +GAGA+CCGGT)	miR-128 M6	r(ACCGGUCUCU <u>G</u> U)
C10	Bt-Sp9-d(AA <u>G</u> +GAGA+CCGGT)	miR-128 M7	r(ACCGGUCUCU <u>C</u> U)

Table S4.3 Purchased oligonucleotides from IDT. Bt= Biotin; Sp9= Spacer C₉H₁₈; scr = scramble control; d = deoxyribonucleotides; r = ribonucleotides.

4.6.2.2 Oligonucleotide Synthesis and Characterization

The LNA/DNA linker was synthesized in 1μmol scale using solid support (CPG 1000Å, Sigma Aldrich) on Expedite Nucleic Acid Synthesis System. Reagents purchased from Sigma Aldrich: TCA Deblock, DCI activator 0.25M, Oxidizer 0.02M, CapA and CapB. Commercial phosphoramidites from Sigma Aldrich (dA(Bz), dC(Bz), dG(ib), dT) were all prepared in 0.07M solutions using dry MeCN. LNA phosphoramidites (indicated with '+' in table below) from

Qiagen (LNA-A(bz), LNA-mC(Bz), LNA-G(dmf)) were added using manually coupling: 25 mg LNA phosphoramidite was dissolved with 400µl dry MeCN (+C: 200µl DCM+200µl MeCN) and 600µl DCI Activator in a 1ml syringe, mixed and added to the column with 50µl/min. The oligonucleotide was cleaved from solid support using 28-30% ammonia solution from Sigma Aldrich at 55°C for 12h. The identity of the oligonucleotide was established by mass spectrometry (MS) using an Autoflex speed MALDI-TOF mass spectrometer (Bruker Daltonics, Hamburg, Germany). Samples were co-spotted with 3-Hydroxypicolinic acid as matrix on a MTP AnchorChip target plate for the analysis. The obtained mass spectra were recorded by flexControl 3.4 (Bruker Daltonics, Germany) software. The oligonucleotide was purified on an Ultimate 3000 UHPLC (Dionex, Sunnyvale, CA, USA) using a DNA-Pac RP (Thermo Fisher Scientific, Waltham, MA, USA) column (4µm, 3.0 × 100 mm²) with a gradient of 5–25% buffer B in A over 30min at 60°C. Buffer A: 0.05M TEAA; Buffer B: 25% buffer A in MeCN. Peaks were monitored at 260nm. The final pure oligonucleotide (87.6% pure) had a yield of 8% (80.6nmol), MS calc. 9487.2 g/mol, MS obs. 9487.106 m/z. HPLC and MALDI spectra are shown below.

Name	Sequence 5' → 3'	MS calc. [g/mol]	MS obs. [m/z]
Linker	T+C+A+CT+GT+GATTTTGAT+GGG+AATAC+CAGACC	9487.2	9487.106

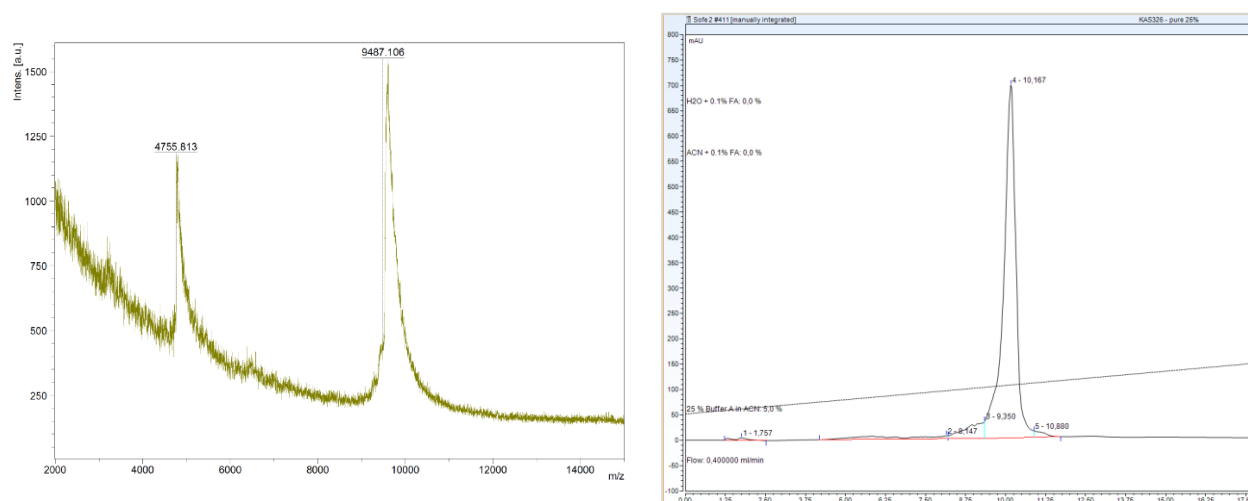


Figure S4.2: MALDI-TOF and UHPLC spectra of purified CTD binding linker.

4.6.3 UV Melting Studies

#	Capture sequence, 5'→3'	Mutation: Position/type	RNA model complement code	RNA complement sequence, 5'→3'
C1	d(AAAGAG+AC+CG+GT)	None, wt	miR128 M1	r(ACCGGUCUCUUU)
C2	d(TAAGAG+AC+CG+GT)	3' terminal, U>A	miR128 M2	r(ACCGGUCUCUUA)
C3	d(CAAGAG+AC+CG+GT)	3' terminal, U>G	miR128 M3	r(ACCGGUCUCUUG)
C4	d(GAAGAG+AC+CG+GT)	3' terminal, U>C	miR128 M4	r(ACCGGUCUCUUC)

4 Mutations in microRNA-128-2-3p Identified with Amplification-Free Hybridization Assay

C5	d(ATAGA+GACC+G+GT)	n-1, U>A	miR128 M5	r(ACCGGUCUCUAU)
C6	d(ACAGA+GACC+G+GT)	n-1, U>G	miR128 M6	r(ACCGGUCUCUGU)
C7	d(+AGA+GAGA+CCGGT)	n-1, U>C	miR128 M7	r(ACCGGUCUCUCU)

Table S4.4 Sequences of model miRNA targets used in UV thermal denaturation study.

T_m given in **Table S4.5** is average of two independent measurements. Representative curves are shown in **Figure S4.2**.

Capture probe #/Target #	miR-128 M1	M2	M3	M4	M5	M6	M7	ΔT_m^{**}
C1 wt	62	39	Nct	Nct	38	Nct	57	17,3
C2	44	61	Nct	43	39	40	57	16,4
C3	45	49	61	50	Nct	40	57	12,8
C4	55	40	Nct	55	47	44	56	6,6
C5	54	41	Nct	38	61	37	Nct	18,5
C6	46	41	45	39	42	55	56	10,2
C7	40	Nct	35	38	41	Nct	64	25,5

Table S4.5: T_m results for duplexes of biotinylated capture probes C1-C7 with complementary and mismatched model miRNA targets. Nct = no clear transition. ** Average difference in T_m for match vs mismatch for duplexes. Only values from clear transition duplexes were used for the calculation.

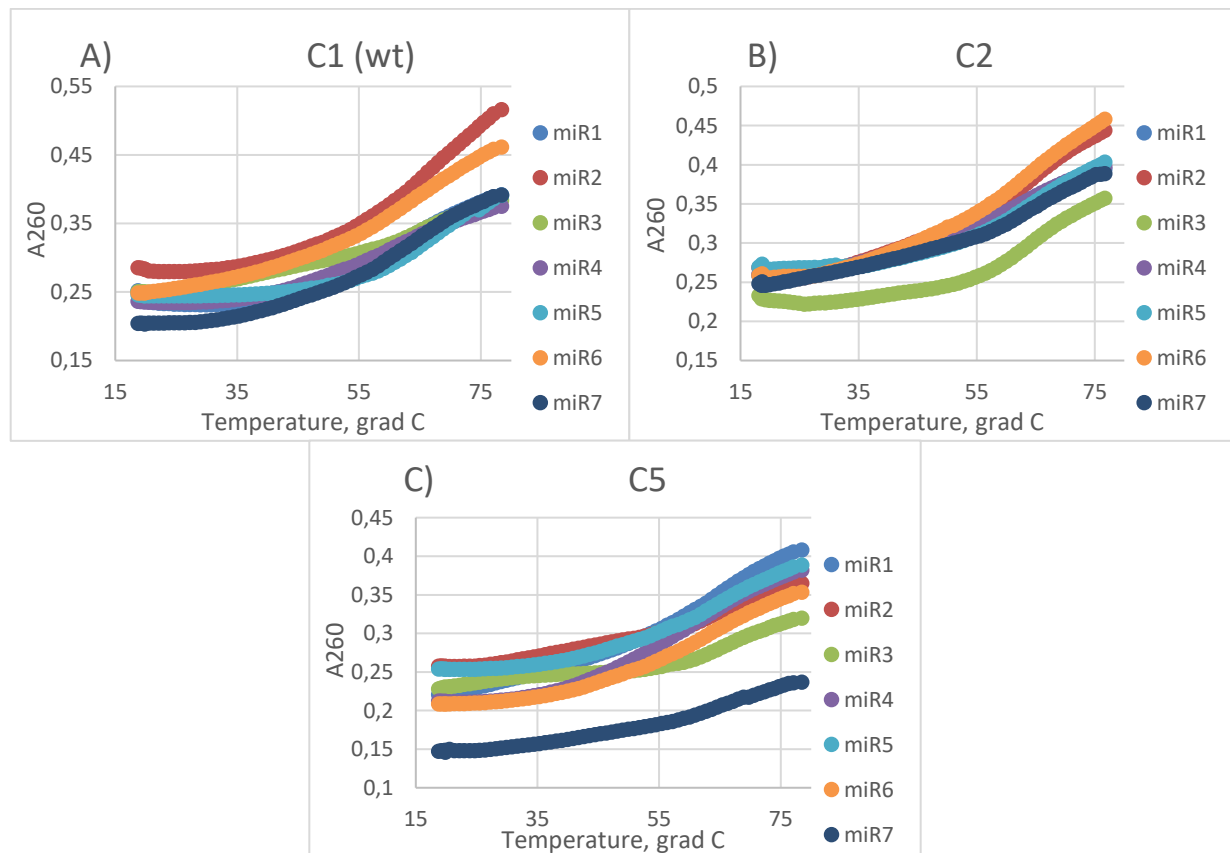


Figure S4.2: Representative UV melting curves for capture probes: A) C1 (wild type), B) C2 (3'-terminal mismatch) and C) C5 (3'-1 mismatch).

4.6.4 Dye study in chip

4.6.4.1 Chip Design and Production

For the fluorophore study, we used microfluidic chips from another related project. The microfluidic chips were made of poly(methyl methacrylate) (PMMA) bonded together with PSA (Pressure-sensitive adhesive) constructed with an inlet hole in first layer, main chamber in second and outlet hole in third layer (**Figure S4.3A**) and measure $1.7 \times 14 \times 41 \text{ mm}^3$ (**Figure S4.3B**). The microfluidic chips were fabricated using laser ablation in 3-layers of 0.5 mm PMMA bonded with 0.1mm PSA engraved with CO₂-laser (Epilog Mini 18 30-watt, Laser cutter). CorelDraw software was applied for drawings for the laser cutter. Prior to cutting, the thin layer of protection plastic was removed from the PMMA surface and the PMMA piece was placed in the epilog on a 1cm plastic piece to reach the right height for cutting. The outlet - and inlet layer were engraved from drawings with the settings: 2 x scans, power 30%, speed 20%, 600 DPI. The middle-layer was added 0.1mm PSA on both sides and engraved with 3 x scans, 20% power, 20% speed and 600 DPI. The 3 layers of PMMA were combined, aligned and bonded together (20kN for 1 min), and broken into smaller chips with individual size $1.7 \times 14 \times 41 \text{ mm}^3$ and chamber volume of $40 \mu\text{l}$ as shown in **Figure S4.3B**.

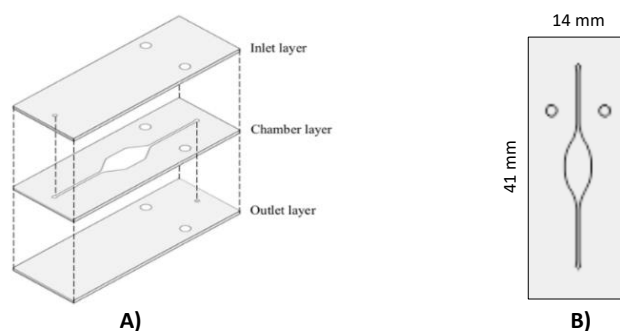


Figure S4.3: One-chamber microfluidic chip. A) Individual layer in the chip aligned. B) One-chamber chipmetrics; $1.7 \times 14 \times 41 \text{ mm}^3$ and chamber volume of $40 \mu\text{l}$.

4.6.4.2 Fluorophore Study in Chip

The average fluorescent emission of the duplicates of each dye (EvaGreen (EG), AccuClear (AC), QuantiFluor (QF), Acridine Orange (AO) and Thiazole Orange (TO)) in different concentrations and in the presence of the CTD (booster) were plotted and shown in **Figure S4.4(A-E)**. **Table S4.6** shows the maximum emission of each dye in concentrations: 1.0%, 2.5%, 3.0% and 4.0%.

Fluorescent Emission (max)					
Dye conc. %	EG	AC	QF	AO	TO
1.0	3,11E+04	1,53E+05	2,01E+05	2,96E+05	5,90E+05
2.5	5,30E+04	3,47E+05	4,65E+05	3,37E+04	3,97E+05
3.0	7,95E+04	2,91E+05	6,22E+05	4,82E+05	3,77E+05
4.0	8,05E+04	4,96E+05	4,77E+05	7,14E+05	2,40E+06
\bar{X}	6,10E+04	3,22E+05	4,41E+05	3,82E+05	9,41E+05

Table S4.6: Average emission (max) of the duplicates of EG (525nm), AC (495nm), QF (525nm), AO (525nm) and TO (529nm) in comparable dye concentrations. Mean maximum emission (\bar{X}) of the different concentrations was also calculated, showing QF and TO giving the overall highest emission.

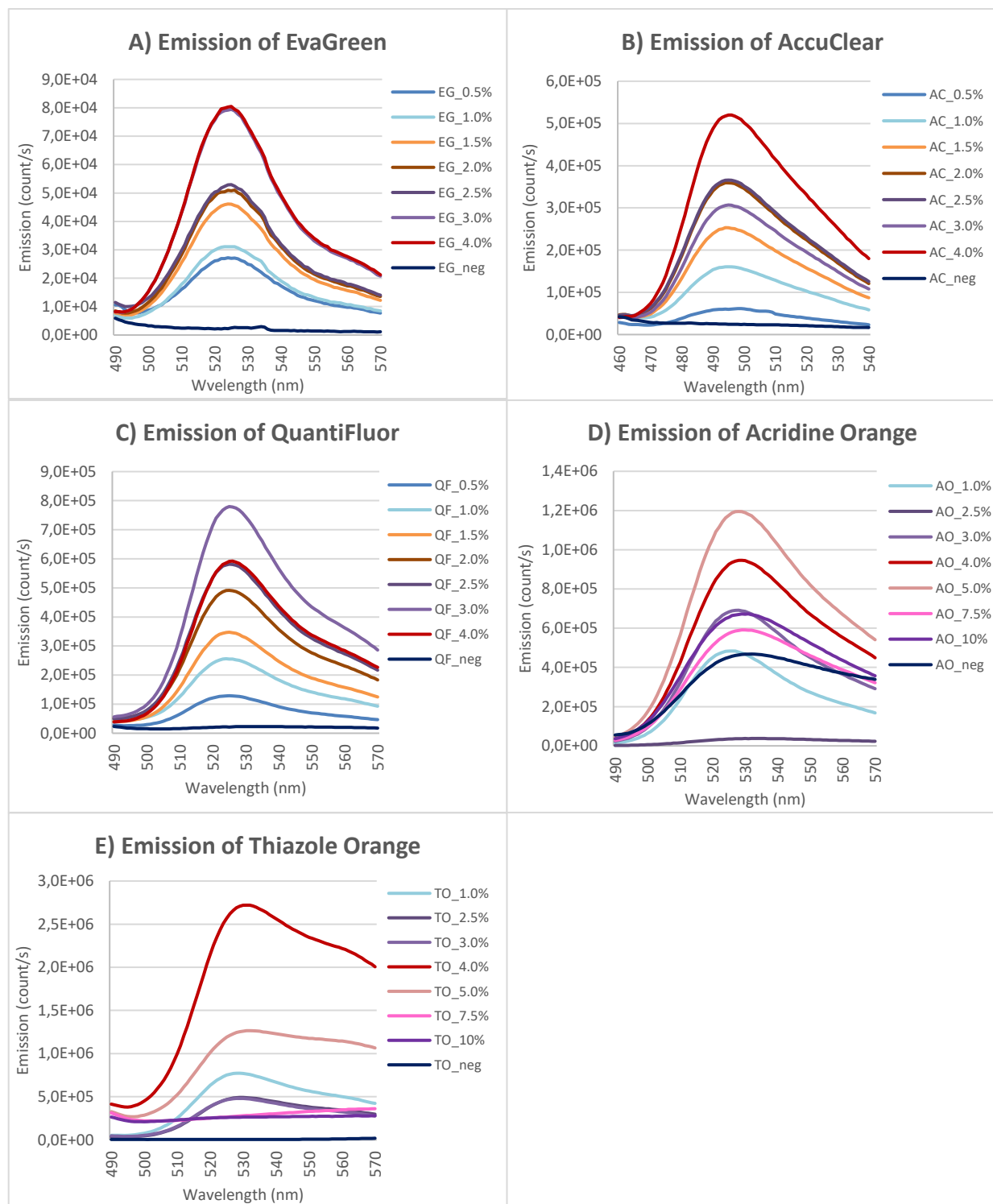


Figure S4.4: Fluorescence emission spectra of five different fluorophores in different conc. in 2 μ M CTD. A) EG, excitation (ex): 480nm B) AC, ex: 450nm C) QF, ex: 480nm D) AO, ex: 480nm and E) TO, ex: 480nm.

4.6.4.3 Photobleaching Study

Photobleaching was calculated in percentage by the formula below. The mean value (\bar{X}) of % photobleaching for each dye in comparable concentrations (1.0%, 2.5%, 3.0%, 4.0%) are shown in **Table S4.7**. Average max emissions for each dye at 1% concentration are plotted in **Figure S4.5**.

$$100 - \left(\frac{\text{max last emission}}{\text{max first emission}} \right) \cdot 100$$

Photobleaching %					
Dye conc. %	EG	AC	QF	AO	TO
1.0	23.5	19.0	2.7	4,1	7,6
2.5	27.4	6.9	3.6	-1,5*	0,3
3.0	25.9	5.4	1.3	22,9	6,1
4.0	8.3	7.2	-1.3*	-2,9*	45,7
\bar{X}	21.3	9.6	2.5	13.5	14.9

Table S4.7: The photobleaching of each dye and its corresponding dye concentrations. *Negative values were not part of the mean calculation.

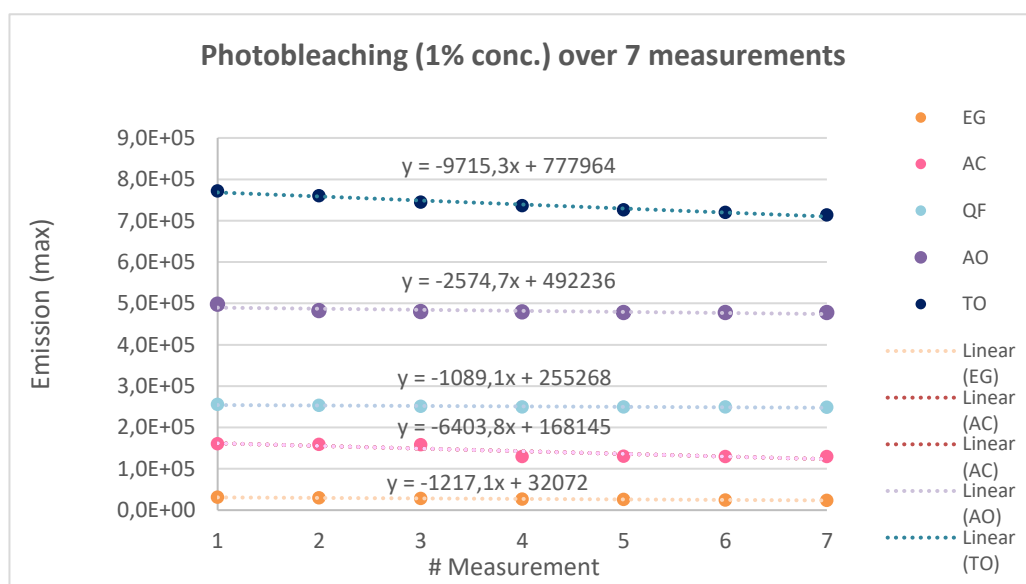


Figure S4.5: Plotted average max emissions for each dye (1%) over 7 repeated measurements.

Maximum (EG, AO, QF) = 525nm; maximum(TO) = 529nm; maximum(AC) = 495 nm.

From **Table S4.7** and **Figure S4.5** it is seen from the mean values of photobleaching [%] over different comparable concentrations and from the slope of the plotted graphs that QF has the overall lowest photobleaching. Negative values were not considered in the calculations. Together with the results from the emission study, QF was chosen as the nucleic acid dye for the miRNA detection bead based assay on its emission qualities and photostability.

4.6.4.4 LOD Study of QuantiFluor

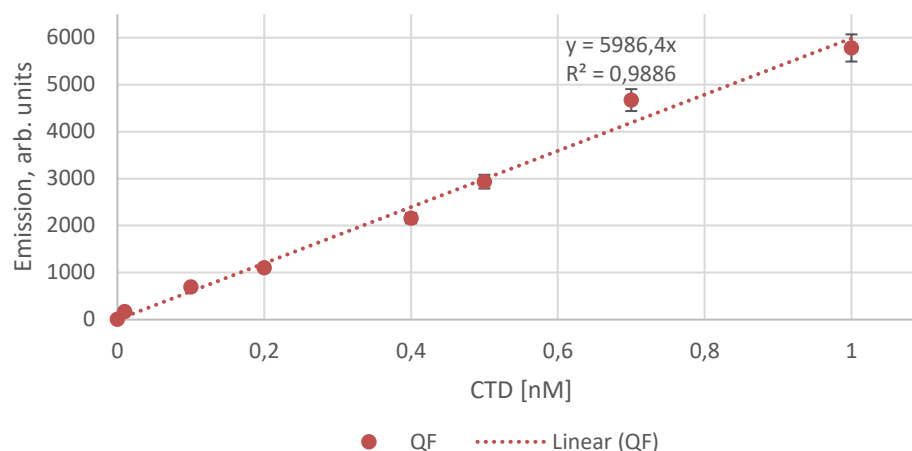


Figure S4.6: Calibration curves for CTD using QuantiFluor dye. The LOD with signal to noise ratio > 3, was determined to be 2.2 pM for QuantiFluor, based on the calibration curves.

4.6.5 Patient Sample Information

Table S4.8: Clinical and demographical information of CRC patients. De-identified human colon cancer samples, along with 10 healthy controls (gender: 5 male and 5 female; Median age at sample (range) 44 (30-53)) were provided by Danish Biobank Regionernes Bio- og GenomBank (RBGB) Herlev, under National Videnskabetisk Komité (National Research Ethics Committee) (NVK) number 21015520.

Pati ent ID	Dise ase type	Age (years)	Gen der	Stage	History of UC/CD*	UC/ CD	Hemoglob in (Hgb)g/dL	ESR (mm/ hour)	Race	Ethnicity
15	CRC	30	M	1	0		6,2	66	Asian	Non Hispanic/ Non-Latino
23	CRC	32	M	1	0		7,4	25	White	Non Hispanic/ Non-Latino
41	CRC	54	M	1	1	UC	8,8	11	White	Non Hispanic/ Non-Latino
24	CRC	61	F	1	0		11	9	NK	NK
54	CRC	45	M	1	1	UC	9,5	14	White	Non Hispanic/ Non-Latino
66	CRC	40	M	2	1	CD	7,3	54	White	Non Hispanic/ Non-Latino
73	CRC	37	F	1	0		8,2	13	White	Non Hispanic/ Non-Latino
111	CRC	62	F	2	0		7	80	White	Non Hispanic/ Non-Latino
122	CRC	48	F	1	0		11,3	66	White	Non Hispanic/ Non-Latino
132	CRC	52	M	1	1	UC	12,5	43	White	Non Hispanic/ Non-Latino
140	CRC	35	M	remission	0		15,4	4	White	Non Hispanic/ Non-Latino
118	CRC	31	M	2	1	UC	11,3	66	White	Non Hispanic/ Non-Latino
156	CRC	44	M	1	1	UC	10,6	50	White	Non Hispanic/ Non-Latino
189	CRC	52	M	remission	0		12	32	NK	NK
221	CRC	38	F	remission	0		14,6	11	White	Non Hispanic/ Non-Latino
212	CRC	50	M	remission	0		17,8	2	White	Non Hispanic/ Non-Latino

4 Mutations in microRNA-128-2-3p Identified with Amplification-Free Hybridization Assay

234	CRC	34	F	1	0		11	11	Asian	Non Hispanic/ Non-Latino
356	CRC	69	M	1	0		8,4	23	White	Non Hispanic/ Non-Latino
401	CRC	60	M	1	1	UC	7,2	60	White	Non Hispanic/ Non-Latino
408	CRC	54	M	1	1	CD	9,9	49	White	Non Hispanic/ Non-Latino

Abbreviations: CRC: Colorectal cancer; *UC: Ulcerative colitis/CD: Crohn's disease; 0: No UC/DC, 1: Yes UC/DC, NK: not known; F: female; M: male.

Table S4.9: Clinical and demographical information of inflammatory bowel disease patients. Deidentified human colitis plasma samples, along with 10 healthy controls (gender: 5 male and 5 female; Median age at sample 44), were obtained from Stanford University (USA) under Stanford Institutional Review Board (IRB protocol 28427).

Patient ID	Disease type	Age (Years)	Gender	Diagnosis	Disease Activity	Hemoglobin (Hgb) g/dL	ESR (mm/hour)	Race	Ethnicity
124	CD	23	M	Crohn's colitis	Flare	7,8	6	NI	Non Hispanic/Non-Latino
131	CD	30	F	Crohn's ileocolitis	Flare	12	2	White	Non Hispanic/Non-Latino
137	CD	35	F	Crohn's ileitis	Remission	11,8	3	African-American	Non Hispanic/ Non-Latino
146	CD	51	F	Crohn's colitis	Remission	12,5	35	White	Non Hispanic/ Non-Latino
150	CD	25	F	Crohn's ileocolitis	Remission	11	17	White	Non Hispanic/ Non-Latino
158	CD	42	M	Crohn's ileocolitis	Flare	12,3	41	White	Non Hispanic/ Non-Latino
159	CD	31	M	Crohn's ileocolitis	Flare	10,6	58	Hispanic	Non Hispanic/ Non-Latino
164	CD	63	M	Crohn's ileocolitis	Remission	13,1	19	White	Non Hispanic/ Non-Latino
171	CD	41	F	Crohn's colitis	Remission	12,9	15	African-American	Non Hispanic/ Non-Latino
174	CD	31	M	Crohn's colitis	Remission	14,7	6	White	Non Hispanic/ Non-Latino
185	CD	57	F	Crohn's ileitis	Flare	9,6	72	NI	Non Hispanic/ Non-Latino
200	CD	68	F	Crohn's colitis	Flare	9,6	35	White	Non Hispanic/ Non-Latino
121	UC	68	F	UC pancolitis	Flare	9,6	62	White	Non Hispanic/ Non-Latino
122	UC	36	F	UC pancolitis	Remission	11,5	70	White	Non Hispanic/ Non-Latino
123	UC	60	M	UC pancolitis	Flare	10,5	65	White	Non Hispanic/ Non-Latino
133	UC	50	M	UC pancolitis	Remission	13,7	9	Asian	Non Hispanic/ Non-Latino
139	UC	37	F	UC pancolitis	Remission	13	9	White	Non Hispanic/ Non-Latino
140	UC	75	F	UC pancolitis	Remission	12,3	19	White	Non Hispanic/ Non-Latino
141	UC	23	F	Ulcerative proctitis	Remission	12,6	6	White	Non Hispanic/ Non-Latino
145	UC	37	F	UC rectosigmoid	Flare	12,3	19	NI	Non Hispanic/ Non-Latino

147	UC	26	M	UC pancolitis	Remission	14,4	6	NI	Hispanic/Latino
186	UC	28	M	UC pancolitis	Flare	8,3	39	Patient declined	Non Hispanic/ Non-Latino
191	UC	19	F	UC pancolitis	Flare	7,6	34	White	Non Hispanic/ Non-Latino
201	UC	60	F	UC pancolitis	Flare	12,8	None	White	Non Hispanic/ Non-Latino

Abbreviations: CD: Crohn's disease; UC: Ulcerative colitis; F: female; M: male; NI: Not Identified

4.6.6 miRNA Analysis

The Ct values for obtained from TaqMan RT-qPCR detected miRNA (wt, 3' and 3'-1 variants) were converted to concentrations by use of the standard curve in **Figure S4.7** and the readouts are shown in **Table S4.14 –S4.16**. **Table S4.13** shows the mean CV% for commercial and in house RT-qPCR of miR128-2-3p and its mutated variants in position 3' and 3'-1 for all patient groups.

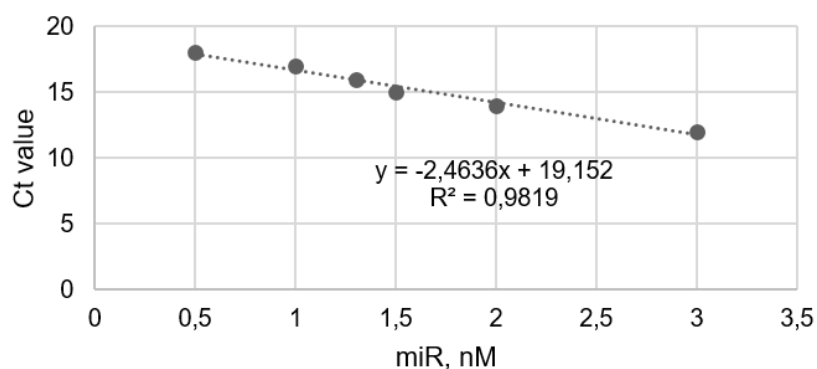


Figure S4.7: Calibration curve for RT-qPCR for wild type analyses (miR128-2-3p).

Table S4.10: Concentrations of miR128 and its variants detected in sera of patients with CRC using amplification-free bead assay. QF standard curve applied to calculate fluorescence signal to concentration. Each data point is an average of two inter-plate replicates; CV average 2.4%.

pat #/miR128, nM, C#	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
15	1,24	0,33	0,34	0,46	0,83	0,41	0,55	0,30	0,56	0,44	0,33	0,31	0,42	0,01
23	0,88	0,42	0,67	0,68	0,53	0,47	0,29	0,45	0,81	0,36	0,47	0,50	0,38	0,01
41	1,45	0,58	0,60	0,40	0,36	0,48	0,72	0,41	0,56	0,73	0,28	0,20	0,36	0,01
24	1,48	0,44	0,30	0,36	0,66	0,57	0,66	0,47	0,83	0,36	0,27	0,24	0,22	0,01
54	0,55	0,50	0,82	0,63	0,70	0,71	0,69	0,33	0,32	0,46	0,23	0,50	0,32	0,01
66	0,28	0,30	0,30	0,32	0,73	0,33	0,36	0,52	0,66	0,36	0,31	0,41	0,24	0,00
73	1,07	0,55	0,46	0,41	0,81	0,82	0,69	0,36	0,39	0,71	0,44	0,20	0,46	0,01
111	0,59	0,31	0,50	0,36	0,48	0,70	0,67	0,42	0,56	0,62	0,35	0,28	0,25	0,01

4 Mutations in microRNA-128-2-3p Identified with Amplification-Free Hybridization Assay

122	0,55	0,57	0,70	0,65	0,50	0,32	0,80	0,43	0,58	0,32	0,28	0,36	0,30	0,01
132	0,74	0,53	0,67	0,60	0,65	0,75	0,61	0,80	0,37	0,79	0,52	0,19	0,22	0,01
140	0,33	0,37	0,32	0,82	0,67	0,73	0,54	0,38	0,80	0,79	0,44	0,45	0,20	0,02
118	0,28	0,41	0,57	0,69	0,71	0,37	0,47	0,43	0,79	0,57	0,52	0,32	0,26	0,01
156	0,81	0,34	0,70	0,43	0,59	0,66	0,64	0,76	0,64	0,31	0,53	0,41	0,15	0,01
189	0,82	0,54	0,63	0,36	0,63	0,65	0,50	0,35	0,53	0,61	0,40	0,19	0,17	0,00
221	1,63	0,60	0,47	0,72	0,35	0,53	0,72	0,80	0,58	0,80	0,35	0,18	0,35	0,02
212	0,96	0,32	0,70	0,47	0,57	0,39	0,47	0,77	0,87	0,47	0,55	0,27	0,20	0,01
234	0,66	0,52	0,41	0,39	0,38	0,76	0,86	0,32	0,59	0,82	0,30	0,36	0,21	0,01
356	1,06	0,62	0,71	0,63	0,29	0,74	0,88	0,52	0,78	0,55	0,38	0,16	0,47	0,01
401	1,39	0,35	0,67	0,29	0,78	0,32	0,41	0,81	0,60	0,35	0,48	0,36	0,18	0,01
408	1,11	0,58	0,37	0,66	0,52	0,49	0,40	0,52	0,58	0,63	0,17	0,23	0,27	0,00

Table S4.11: Concentrations of miR128 and its variants detected in sera of patients with colitis using amplification-free bead assay. Each data point is an average of two inter-plate replicates; CV average 2.2%.

pat	#/miR128,	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
nM, C#															
124	1,14	0,62	0,61	0,39	0,72	0,46	0,49	0,71	0,66	0,64	0,44	0,44	0,16	0,01	
131	0,97	0,4	0,8	0,37	0,47	0,48	0,48	0,57	0,66	0,45	0,46	0,19	0,36	0,01	
137	0,49	0,32	0,44	0,57	0,39	0,71	0,51	0,54	0,81	0,72	0,25	0,47	0,4	0,02	
146	1,45	0,57	0,32	0,42	0,65	0,43	0,43	0,77	0,57	0,41	0,15	0,39	0,17	0	
150	0,82	0,58	0,4	0,6	0,42	0,42	0,76	0,51	0,79	0,75	0,48	0,36	0,22	0,01	
158	0,22	0,32	0,37	0,54	0,31	0,62	0,52	0,58	0,34	0,8	0,42	0,42	0,47	0,02	
159	0,93	0,51	0,31	0,49	0,57	0,68	0,84	0,62	0,65	0,37	0,28	0,19	0,2	0,01	
164	0,72	0,54	0,41	0,66	0,55	0,64	0,39	0,64	0,58	0,45	0,24	0,19	0,3	0,01	
171	0,4	0,4	0,6	0,63	0,5	0,68	0,7	0,74	0,78	0,8	0,43	0,43	0,48	0,01	
174	0,64	0,41	0,7	0,74	0,83	0,67	0,71	0,55	0,78	0,8	0,16	0,25	0,19	0,01	
185	0,38	0,53	0,45	0,35	0,82	0,37	0,31	0,65	0,81	0,67	0,2	0,46	0,31	0,01	
200	0,22	0,29	0,61	0,66	0,41	0,42	0,5	0,61	0,54	0,47	0,4	0,27	0,3	0,02	
121	0,84	0,33	0,5	0,59	0,73	0,4	0,38	0,64	0,64	0,6	0,5	0,27	0,36	0,02	
122	1,47	0,39	0,53	0,64	0,62	0,37	0,8	0,43	0,32	0,34	0,51	0,17	0,29	0,01	
123	0,97	0,49	0,35	0,51	0,49	0,44	0,5	0,52	0,71	0,78	0,33	0,41	0,25	0,01	
133	1,06	0,6	0,74	0,54	0,63	0,55	0,72	0,78	0,59	0,81	0,24	0,21	0,35	0,01	
139	0,97	0,51	0,65	0,67	0,57	0,34	0,39	0,38	0,72	0,43	0,45	0,49	0,32	0,02	
140	0,88	0,57	0,71	0,39	0,34	0,6	0,64	0,63	0,36	0,35	0,43	0,24	0,32	0	
141	1,31	0,39	0,31	0,4	0,79	0,74	0,77	0,41	0,71	0,61	0,47	0,14	0,37	0,02	
145	1,64	0,39	0,31	0,49	0,39	0,45	0,31	0,47	0,32	0,32	0,26	0,39	0,35	0,02	
147	0,31	0,42	0,35	0,71	0,27	0,73	0,69	0,49	0,47	0,79	0,22	0,46	0,4	0,01	
186	0,15	0,6	0,72	0,79	0,48	0,61	0,77	0,7	0,54	0,58	0,42	0,39	0,44	0,01	
191	0,34	0,46	0,66	0,77	0,45	0,77	0,34	0,88	0,71	0,42	0,44	0,39	0,16	0,01	
201	0,51	0,55	0,6	0,79	0,71	0,31	0,52	0,72	0,73	0,42	0,47	0,21	0,49	0,01	

Table S4.12: Concentrations of miR128 and its variants detected in sera of healthy controls using amplification-free bead assay. Each data point is an average of two inter-plate replicates; CV average 2.2%.

pat #/miR128, nM, C#	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
1	0,36	0,40	0,27	0,41	0,36	0,50	0,45	0,18	0,32	0,38	0,36	0,25	0,30	0,00
2	0,23	0,41	0,29	0,16	0,21	0,31	0,47	0,30	0,44	0,36	0,21	0,32	0,17	0,00
3	0,23	0,34	0,25	0,22	0,27	0,22	0,37	0,43	0,33	0,39	0,24	0,18	0,37	0,00
4	0,54	0,43	0,48	0,42	0,46	0,37	0,42	0,23	0,26	0,23	0,31	0,46	0,33	0,00
5	0,14	0,58	0,42	0,45	0,40	0,28	0,20	0,31	0,20	0,16	0,32	0,50	0,50	0,00
6	0,07	0,56	0,42	0,48	0,19	0,20	0,28	0,21	0,26	0,32	0,32	0,36	0,27	0,00
7	0,36	0,37	0,48	0,17	0,43	0,25	0,45	0,18	0,24	0,19	0,42	0,32	0,38	0,00
8	0,16	0,38	0,26	0,19	0,46	0,47	0,29	0,36	0,46	0,49	0,50	0,45	0,47	0,00
9	0,14	0,40	0,44	0,37	0,43	0,43	0,43	0,35	0,44	0,33	0,17	0,38	0,32	0,00
10	0,22	0,49	0,51	0,32	0,36	0,22	0,48	0,51	0,46	0,40	0,22	0,24	0,29	0,00
11	0,10	0,65	0,31	0,22	0,30	0,14	0,50	0,49	0,24	0,23	0,45	0,33	0,22	0,00
12	0,08	0,33	0,35	0,45	0,19	0,38	0,19	0,52	0,38	0,18	0,23	0,43	0,15	0,00
13	0,27	0,59	0,33	0,44	0,40	0,37	0,46	0,30	0,15	0,29	0,21	0,44	0,38	0,00
14	0,23	0,45	0,34	0,51	0,39	0,16	0,22	0,48	0,27	0,44	0,35	0,28	0,38	0,00
15	0,56	0,64	0,29	0,25	0,18	0,17	0,28	0,39	0,17	0,16	0,47	0,24	0,17	0,00
16	0,28	0,31	0,48	0,33	0,32	0,24	0,17	0,33	0,51	0,51	0,14	0,27	0,34	0,00
17	0,22	0,60	0,30	0,27	0,33	0,34	0,27	0,23	0,32	0,27	0,23	0,29	0,23	0,00
18	0,34	0,53	0,37	0,41	0,30	0,42	0,49	0,16	0,35	0,25	0,32	0,15	0,27	0,00
19	0,42	0,61	0,22	0,49	0,44	0,31	0,34	0,16	0,36	0,31	0,27	0,28	0,14	0,00
20	0,37	0,39	0,26	0,36	0,40	0,37	0,17	0,41	0,51	0,50	0,52	0,15	0,33	0,00

Table S4.13: Mean CV% for commercial and in house RT-qPCR of miR128 and its mutated variants in positions 3' and 3'-1.

Patient group	1 (wt)	2 (3', U>A)	3 (3', U>G)	4 (3', U>C)	5 (3'-1, U>A)	6 (3'-1, U>G)	7 (3'-1, U>C)
CRC	3,4	2,3	2,7	2,9	2,8	2,7	2,9
Colitis	3,3	3,6	5,3	6,2	7,0	8,1	6,5
Healthy	4,5	3,1	3,9	3,4	3,2	4,4	5,6

Table S4.14: Concentrations of miR128 and its variants detected in sera of patients with CRC using qPCR. Each data point is an average of two inter-plate replicates; CV average 2.8%.

pat #/miR128, nM, mut variant#	1	2	3	4	5	6	7
15	0,83	1,52	1,57	1,65	1,97	1,61	1,73
23	1,48	0,83	1,93	1,97	1,85	1,77	1,65
41	0,51	0,14	1,81	1,65	1,61	1,69	1,89
24	2,22	1,08	1,81	1,89	2,13	2,05	2,13
54	0,96	1,24	1,93	1,77	1,81	1,81	1,81
66	0,39	1,04	1,61	1,65	1,97	1,65	1,69
73	0,87	0,35	1,12	1,08	1,44	1,44	1,32
111	1,20	0,96	1,89	1,81	1,89	2,09	2,05
122	0,75	0,87	1,69	1,65	1,52	1,36	1,77
132	1,40	1,28	2,09	2,05	2,09	2,17	2,05
140	0,67	1,16	1,48	1,93	1,77	1,85	1,69
118	0,55	0,96	1,85	1,97	1,97	1,69	1,77
156	0,59	0,83	1,89	1,65	1,81	1,85	1,85
189	0,14	-0,18	2,09	1,89	2,09	2,09	1,97
221	1,93	0,30	1,48	1,69	1,40	1,52	1,69
212	1,69	1,00	1,97	1,77	1,85	1,73	1,77
234	1,00	1,12	1,65	1,65	1,61	1,93	2,01
356	1,69	1,52	2,17	2,09	1,81	2,17	2,30
401	1,69	0,63	1,77	1,48	1,89	1,48	1,57
408	1,57	1,12	1,69	1,93	1,81	1,77	1,69

Table S4.15: Concentrations of miR128 and its variants detected in sera of patients with colitis using qPCR. Each data point is an average of two inter-plate replicates; CV average 5.7%.

pat #/miR128, nM, mut variant#	1	2	3	4	5	6	7
124	0,71	2,17	1,40	1,40	1,32	2,26	2,50
131	1,57	0,51	2,05	1,69	2,91	2,54	2,34
137	0,59	0,35	2,66	2,54	2,91	1,89	2,54
146	2,17	1,16	2,22	2,42	2,09	2,50	3,39
150	1,20	1,24	2,50	2,54	2,50	2,66	1,85
158	0,35	1,24	1,69	2,50	2,66	2,58	1,81
159	0,71	0,43	0,96	1,12	1,20	1,32	1,44

164	1,52	2,13	2,01	2,54	2,13	2,38	2,50
171	0,63	0,83	1,61	1,65	2,91	0,06	1,20
174	1,32	1,40	2,13	2,38	2,87	2,38	2,62
185	0,71	0,71	1,61	1,52	1,93	2,05	2,26
200	0,51	0,96	2,38	2,99	2,30	2,50	2,34
121	0,63	1,00	2,30	2,17	2,62	2,30	2,05
122	0,91	0,10	2,01	2,09	2,09	1,89	2,22
123	1,32	0,06	1,40	1,93	2,13	2,09	1,52
133	1,77	1,04	2,01	1,85	2,34	2,50	1,97
139	1,28	1,36	1,57	1,36	2,34	1,61	1,65
140	1,85	1,32	2,17	1,89	2,22	2,54	2,09
141	2,26	1,89	2,05	1,57	1,89	1,85	1,89
145	2,09	1,32	2,26	2,54	2,46	2,54	2,38
147	0,67	0,26	1,40	1,69	1,32	1,73	1,69
186	0,96	1,28	1,97	2,05	1,81	2,54	2,42
191	0,67	1,44	1,04	2,54	1,69	2,83	2,78
201	1,20	0,91	2,50	2,22	2,38	1,85	1,12

Table S4.16: Concentrations of miR128 and its variants detected in sera of healthy controls using qPCR.

Each data point is an average of two inter-plate replicates; CV average 4.0%.

pat #/miR128, nM, mut variant#	1	2	3	4	5	6	7
1	0,14	2,46	1,48	1,44	1,57	2,46	1,65
2	0,83	0,51	1,65	1,69	3,31	1,04	1,77
3	0,10	0,10	1,52	1,40	1,52	1,40	1,61
4	1,36	0,91	1,97	1,73	1,97	1,69	1,93
5	0,59	0,55	1,57	1,08	1,57	0,96	1,40
6	0,18	1,32	1,73	1,89	1,24	1,65	1,61
7	0,10	0,55	1,12	1,24	1,89	1,32	1,08
8	0,79	0,91	1,69	1,69	1,61	1,93	1,32
9	0,39	0,71	1,48	1,40	1,73	1,44	1,89
10	0,91	1,04	1,52	1,81	1,00	1,73	1,65
11	0,47	0,59	1,81	1,40	1,73	1,32	1,77
12	0,35	0,75	1,52	1,77	1,24	1,69	0,83
13	0,02	0,87	1,65	1,69	1,89	1,61	1,89
14	0,10	0,10	1,00	2,01	1,89	1,73	1,28
15	0,91	-0,18	1,73	1,32	1,24	1,24	1,77

4 Mutations in microRNA-128-2-3p Identified with Amplification-Free Hybridization Assay

16	1,04	0,83	1,48	1,65	1,65	1,61	1,52
17	0,59	1,00	1,81	1,52	1,57	2,26	2,42
18	1,04	1,36	2,46	2,91	2,26	1,93	2,83
19	0,75	0,83	2,01	2,13	0,87	1,85	2,26
20	2,05	2,91	1,57	2,05	1,69	2,01	2,30

4.6.7 Statistical analyses

Table S4.17: Linear regression study for amplification-free bead assay vs qPCR. Analysis done for all patient groups – Pearson correlation coefficients.

C1/PCR1	C2/PCR2	C3/PCR3	C4/PCR4	C5/PCR5	C6/PCR6	C7/PCR7
0,66	-0,02	0,26	0,47	0,19	0,29	0,10

Table S4.18: Pearson coefficients for clinical correlation of amplification-free assay.

	Age (years)	Gender	Stage	History of UC/CD*	Hemoglobin (Hgb) g/dL	ESR (mm/hour)	Race
Age (years)	1						
Gender	0,04	1,00					
Stage	0,11	-0,14	1,00				
UCCD history	-0,06	-0,03	-0,05	1,00			
Hg	-0,03	0,06	-0,69	0,03	1,00		
ESR	0,24	-0,08	0,55	0,05	-0,50	1,00	
Race	-0,26	-0,05	-0,08	0,09	-0,04	0,01	1,00
C1	0,09	0,13	-0,21	-0,16	0,01	-0,11	-0,17
C2	0,07	-0,02	-0,18	-0,02	0,01	-0,21	0,02
C3	0,13	-0,12	0,00	0,05	-0,02	-0,13	-0,09
C4	-0,21	-0,08	-0,23	0,09	0,29	-0,15	0,06
C5	-0,06	-0,12	0,12	0,02	-0,14	0,21	-0,10
C6	-0,19	-0,09	-0,17	-0,17	0,08	-0,34	0,26
C7	-0,11	0,03	-0,17	-0,22	0,11	-0,14	0,34
C8	0,15	-0,01	-0,13	0,38	0,16	0,00	0,10
C9	-0,10	0,01	-0,05	-0,15	0,10	-0,16	0,12
C10	-0,15	-0,12	-0,23	-0,02	0,27	-0,26	0,12
C11	-0,10	0,05	0,07	-0,06	-0,02	0,04	-0,15
C12	-0,30	-0,07	0,09	0,14	-0,14	0,03	0,07
C13	0,04	0,16	-0,02	0,00	-0,04	-0,18	0,20

5 MicroRNA Pools Synthesized Using Tandem Solid-Phase Oligonucleotide Synthesis

Sofie Slott^a and Kira Astakhova^a

^a Department of Chemistry, Technical University of Denmark, Kemitorvet 206-207, Kgs. Lyngby, Denmark

Abstract: Herein we describe a new approach to make pools of microRNA targeting breast cancer cells. The microRNA pools were synthesized at once on the same solid-support using ‘Tandem Oligonucleotide Synthesis’ strategy. We make up to four consecutive microRNA (miR129-1-5p, miR31, miR206 and miR27b-3p) using 2’/3’OAc nucleotide phosphoramidites, with the total length of the pool reaching 88 nucleotides. The developed phosphoramidites combined give a cleavable moiety that separates the microRNAs and is cleaved using standard post-RNA synthesis cleavage conditions. Furthermore, we investigate making branched pools (microRNA dendrimers) versus linear pools as a strategy to further improve the product yields. Our approach provides with microRNA pools in high yields which is of relevance to the growing demand on synthetic RNA oligomers for nucleic acid research and technology.

5.1 Tandem Oligonucleotide Synthesis

Solid-phase oligonucleotide synthesis has since its development revolutionized the field of nucleic acid research and applications in molecular diagnostics and gene therapy^{101,163}. Today, solid-phase synthesis is the preferred way of synthesizing short single stranded oligonucleotides. It is applied for making FDA-approved gene therapeutics in big scale such as small interfering RNA (siRNA) and antisense oligonucleotides (ASO)^{40,164,165}.

Many applications require two or more oligonucleotides used at the same time. Therefore, the ability to make multiple oligonucleotides on the same solid-support (*Tandem Oligonucleotide Synthesis (TONS)*)^{166,167} could potentially improve the synthesis and enhance studies of synthetic oligonucleotide pools. Despite the development of microarray based oligonucleotide synthesis, where thousands of oligonucleotides are synthesized parallel in small scale^{105–108}, there are no straightforward solutions for making multiple oligonucleotides at once yet.

TONS was first investigated by Hardy et al. in 1994 describing a cleavable linker reagent, *TOPS* (Two Oligomers Per Synthesis) to make several oligonucleotides on the same solid-support released under basic conditions (**Figure 5.1A,B**)^{167,12}. The oligonucleotides were cleaved from the DNA pool via a cyclic phosphate formation using 40% methylamine in aqueous ammonia at 60°C overnight. To avoid these relatively harsh conditions, Pon et al. prepared a cleavable 3'-O-hydroquinone-*O,O'*-diacetic acid linker arm (Q-linker) (**Figure 5.1C**) that released the oligonucleotides after treatment with ammonia for 2 min¹⁶⁶. Pon et al. used this linker for several applications such making up to six consecutive DNA 20mers (120 in total), various trinucleotide codons, primer pairs for PCR and self-complementary strands for *in situ* formation of double-stranded DNA¹⁰⁴. However, the Q-linker required 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate and 4-Dimethylaminopyridine (HBTU/DMAP) as coupling reagent instead of standard coupling reagents for automated oligonucleotide synthesis. Pon et al. described the idea of using the linker for making complementary siRNA pairs in the future, however this has not been done yet.

In 2021, Yamamoto et al. designed four different cleavable spacers for tandem synthesis of multiple oligonucleotides: two diacetylated derivative linkers and two succinic derivative linkers (**Figure 5.1D**).¹⁶⁸ These were inspired by commercially available 'uni-linkers' on solid-support, which are cleaved easily under mild conditions. They describe that cyclic diols are cleaved faster than acyclic analogues because of the locked conformation. They also use the cleavable linkers to make DNA primer pairs. A different approach was described by Khalil Arar et al. in a filed patent from 2002, having different anchor linkers on the support that were orthogonal protected¹⁶⁹. The first linker are deprotected, followed by standard phosphoramidite chemistry to make first oligonucleotide (ON1) which is capped, and then the second linker is deprotected, and the second ON2 is made¹⁶⁹.

To our knowledge, none of the described linkers has been applied to make pools of RNA or complementary siRNA pairs, but only focus on DNA applications. Furthermore, both Hardy et al. and Pon et al. experienced challenges with their linkers. Even with extremely high coupling efficiency there will be a difference in yield of the product oligonucleotides in the pool: a 20mer will give 1.5:1 oligonucleotide product ration in the pool, and 35mer will lead to 2:1 ratio.¹⁶⁷ As a result, the products are present in the pool in different amounts in the pool, which will affect the biological properties.

In this work, we describe a new method to make miRNA pools using tandem solid-phase synthesis strategy and cleavable sites. We explore both linear and branched pool synthesis strategies using new reagents that we prepared. We confirm that the pools of microRNA can be successfully prepared using our new reagents, and that the yields of the individual microRNAs in the pool are nearly equivalent to each other.

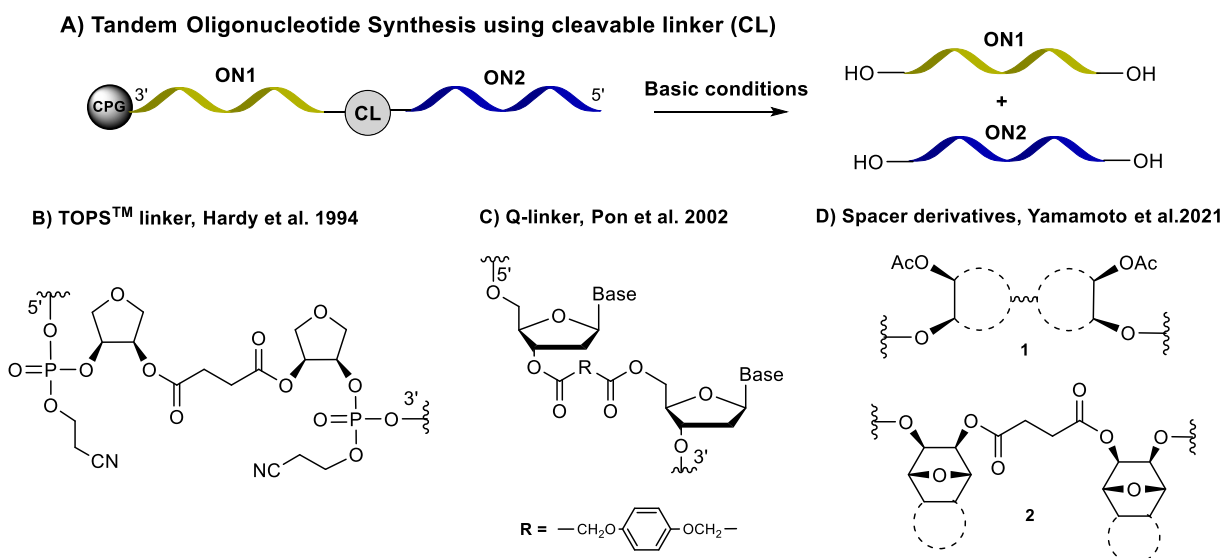


Figure 5.1: Cleavable linkers from literature applied for tandem oligonucleotide synthesis. (A) Principle of Tandem Oligonucleotide Synthesis (TONS) with cleavable linker (CL) in between.; (B) TOPS linker by Hardy et al.¹⁶⁷; (C) Q- linker by Pon et al.¹⁶⁶, and (D) Cleavable Spacer derivatives by Yamamoto et al.: Diacetyl linker derivatives (1) and succinic diester derivatives (2).¹⁶⁸

5.2 Results and discussion

5.2.1 Study design

It is critical when designing the experimental conditions to avoid RNA degradation during work up. According to our design, the product microRNA pool is cleaved easily with basic conditions compatible with processing RNA products. The TONS is used to make six pools: three linear containing 2, 3, and 4 consecutive microRNAs, and three branched pools (**S3** in between **S8** and **S6**) containing 2, 3 and 4 miRNAs with cleavable sites in between each oligonucleotide (**Figure 5.2**). The branched structure was designed from a theoretical perspective to compensate for the yield difference described by Hardy et al. and Pon et al., between the first and last synthesized oligonucleotide using cleavable sites (CS) in TONS (**Figure 5.2B**)^{104,167}. Herein, we use the branching point to investigate the possibility of obtaining higher yield of oligonucleotides in the outer layers generating a dendrimer structure. For this purpose, we synthesized a branched linker (**S3**) that contains a 2-cyanoethyl phosphoramidite in the middle and 4,4'-Dimethoxytrityl (DMTr) groups on each side. Such two nucleotides can react and generate a branched structure (**Figure 5.2**) during automated synthesis. The branched linker molecule (**S3**) has previously been synthesized by Hervé et al. for labeling oligonucleotides with biotin¹⁷⁰.

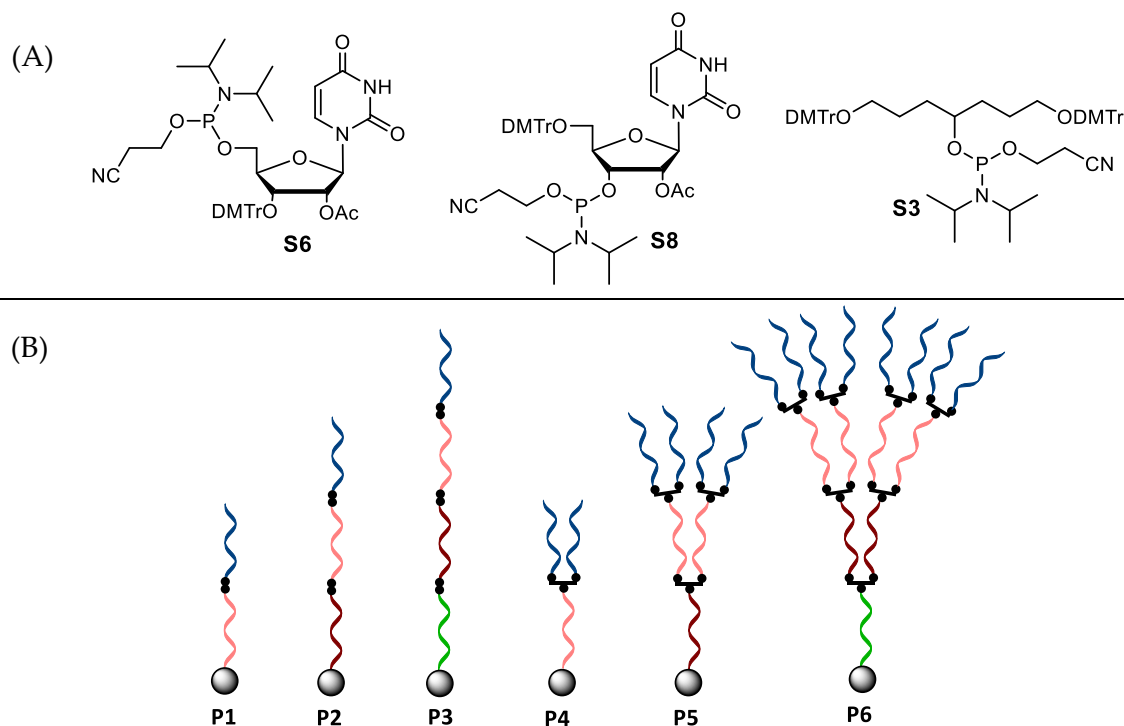


Figure 5.2: (A) Phosphoramidite reagents applied in this study and (B) Pools of microRNA synthesized in this thesis. Linear (P1-P3) and branched (P4-P6) pools of microRNA were synthesized using solid-phase with cleavable sites (CS). S8 and S6 = ● ; S3 = ■

The CS that we use in creating microRNA pools consist of two 2'/3'OAc nucleotide phosphoramidites (S8 and S6 (reverse)) which in combination gives a diacetylated derivative (**Figure 5.3**). S6 and S8 are coupled in the standard automated synthesis, and upon cleavage of product from solid-support followed by deprotection with base, the two acetyl groups are simultaneously removed. Due to the basic conditions, the hydroxyl groups on 2'/3' position will be deprotected and make an intramolecular nucleophilic attack on the phosphate to generate 2',3'-cyclic phosphate dinucleotide and release the two oligonucleotides on each site (**Figure 5.3**). 2'-O-Acetyl RNA phosphoramidites is usually not applied as an alternative to 2'O-TBDMS RNA¹⁷¹ monomers mainly due to ease of 2',3'-migration during monomer synthesis and lack of compatibility with other protecting groups¹⁷¹⁻¹⁷³. However, in our tandem synthesis approach we can use both regioisomers.

Next, we selected four microRNAs (miR31, miR206, miR129-1-5p and miR27-3p) that are reported to be dysregulated in breast cancer (**Figure 5.4, Table S5.1**)¹⁷⁴⁻¹⁷⁷. MicroRNA are small regulatory non-coding RNA with 18-22 nucleotide length. MicroRNA has shown to be dysregulated in many cancers and serve as either tumor-suppressor miRNA (down-regulated) or oncogenic miRNA (up-regulated)³⁹.

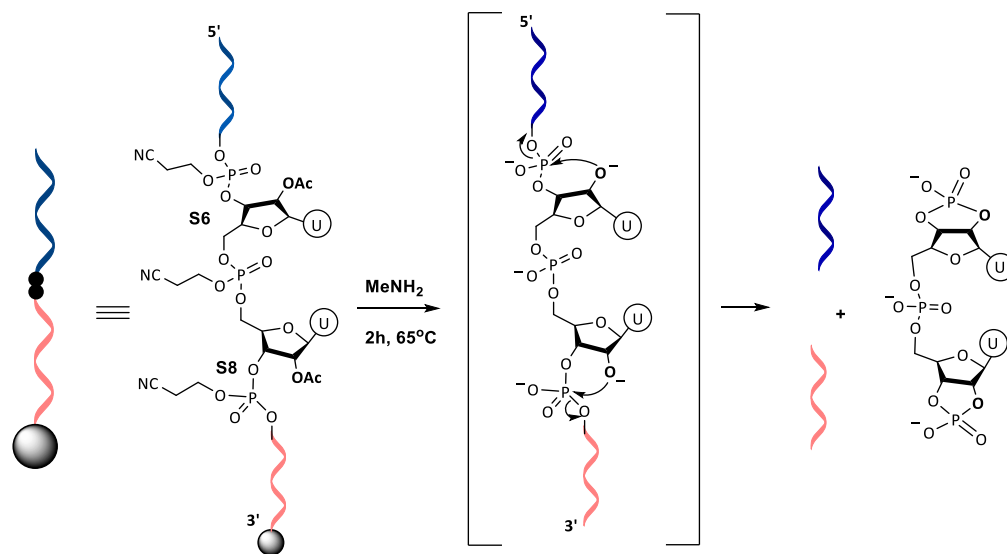


Figure 5.3: Proposed cleavage mechanism of linear CS. Methylamine solution, 2h 65°C to generate RNA pools with 2',3'-cyclic phosphate dinucleotide as byproduct.

MicroRNA therapy is a rapidly developing field where drug candidates can either mimic (miR analogues) or inhibit microRNA function (antimiR)^{39,40}. Furthermore, microRNA therapy has shown relevance when given in combination with other cancer therapies to combat chemo resistance of some drugs¹⁷⁸. Studies have also shown that using the combination of multiple microRNAs increases the therapeutic efficacy^{41,179}.

The microRNA used in this work; hsa-miR129-1-5p, hsa-miR31, hsa-miR206 and hsa-miR27b-3p are all dysregulated in breast cancer^{174–177}. miR129-1-5p, miR31 and miR206 act as tumor-suppressors in breast cancer and are known to be downregulated^{175–177}. Conversely miR27b-3p has recently shown to have an oncogenic function in breast cancer and is upregulated (**Table S5.1**)¹⁷⁴. To increase the stability, we incorporated 2'OMe modifications in terminal and middle positions of miRNAs. The sequences were obtained from miRBase¹⁸⁰, and the correct 3D folding was confirmed with Nupack RNA folding prediction.

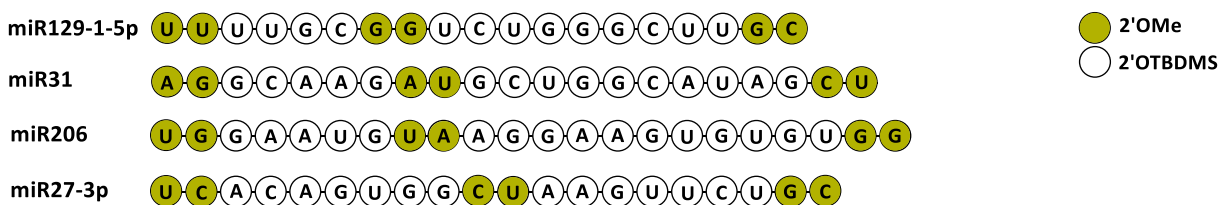
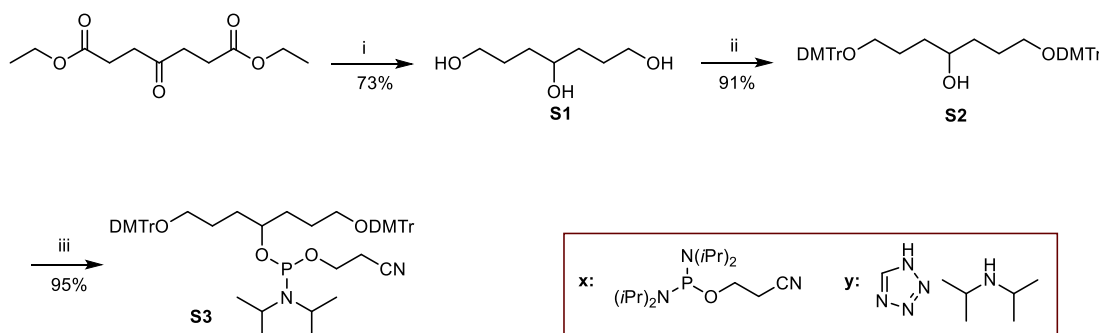


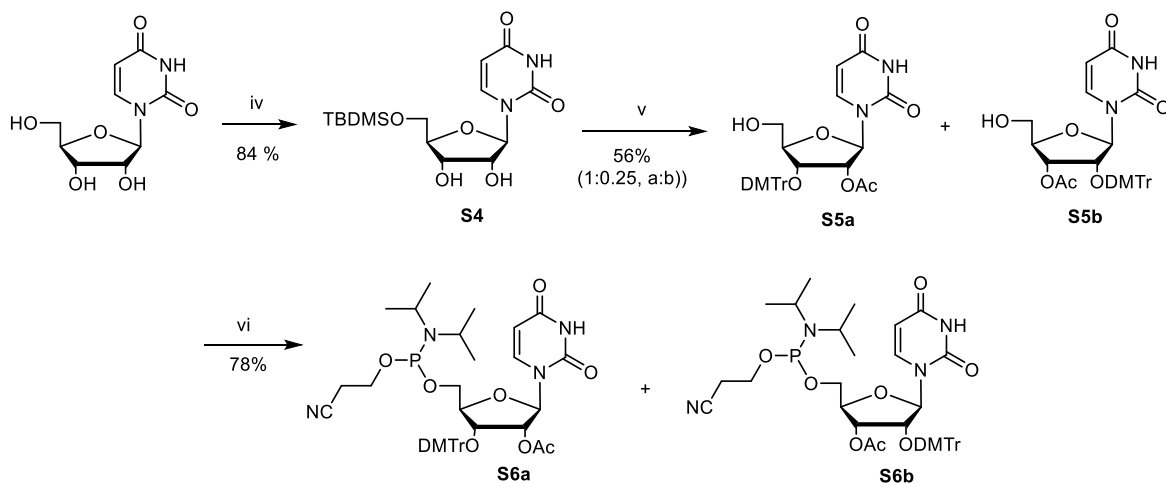
Figure 5.4: Modified microRNA sequences (5'→3') used in this study. miR129-1-5p, miR31, miR206 and miR27-3p. TBDMS = *Tert*-Butyldimethylsilyl; Me = Methyl.

5.2.2 Synthesis of monomers

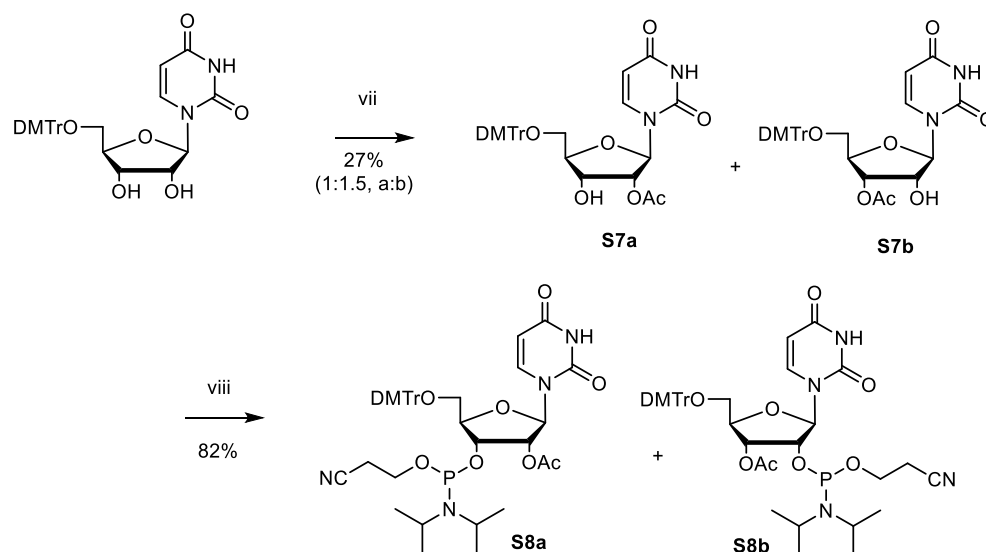
The branched phosphoramidite reagent **S3** was synthesized in three steps starting from diethyl-4-oxoheptanedioate¹⁷⁰ (**Scheme 5.1**). First, the diester was reduced using LiAlH_4 solution affording the triol (**S1**). The primary alcohols were then protected using DMTr chloride under standard conditions to obtain **S2**. The final phosphitylation was done using 2-cyanoethyl *N,N,N',N'*-tetraisopropyl phosphoramidite (x) and diisopropylammonium tetrazolide (y) under dry and inert conditions yielding the final compound 1,7-di-O-dimethoxytrityl heptane-4-O-(2-cyanoethyl)-*N,N*-diisopropylphosphoramidite (**S3**).



Scheme 5.1: Synthesis of branched linker (S3), reagents and conditions: (i) LiAlH_4 , THF, rt., o/n; (ii) DMTr-Cl, Pyridine, rt., o/n; (iii) 2-cyanoethyl *N,N,N',N'*-tetraisopropylphosphoramidite (x), diisopropylammonium tetrazolide (y), DCE, DCM, rt. o/n.



Scheme 5.2: Synthesis of 2'/3'OAc-5'phosphoramidite (S6), reagents and conditions: (iv) TBDMS-Cl, imidazole, DMF, pyridine, rt. 3h; (v) 1) DMTr-Cl, pyridine, rt. 48h 2) Acetic anhydride, NMI, rt. o/n; 3) TBAF, THF, rt., 4h; (vi) as in (iii).



Scheme 5.3: Synthesis of 2'/3'OAc-3'/2'Phosphoramidite (S8**)**, reagents and conditions: (vii) AcCl, pyridine, THF, rt, 3h; (viii) as in (iii).

The first part of our CS is 2'/3'OAc-5'phosphoramidite **S6** (Scheme 5.2). **S6** was synthesized as regioisomeric mixture (**S6a**, **S6b**) in three steps from uridine as described in literature^{181,182}. The primary alcohol of uridine was first protected using TBDMS-Cl to obtain **S4** without further purification. Then a standard DMTr-Cl protection was performed, followed by acetylation with anhydride, and finally removal of TBDMS protecting group using TBAF affording **S5** as a regioisomeric mixture (**S5a** and **S5b**). The final phosphitylation reaction was done using the same conditions as (iii) to yield the final compounds **S6a** on **S6b**. Both regioisomers were applied in the solid-phase RNA synthesis. The second part of our CS is 2'/3'OAc-3'/2'Phosphoramidite (**S8**) (Scheme 5.3). It was prepared in two steps from commercially available 5'-ODMTTr-uridine as described in literature¹⁷¹. The monoacetylation was made using acetyl chloride to get a mixture of **S7a** and **S7b**. The phosphitylation was done using same conditions as in (iii) affording **S8** as a mixture of isomers (**S8a** and **S8b**). All three synthesized phosphoramidites (**S3**, **S6** and **S8**) were used directly in automated synthesis of the pools (Table 5.1).

5.2.3 Synthesis of miRNA pools

By applying standard 2'-OTBDMS-phosphoramidites, 2'-OMe-phosphoramidites, and the synthesized phosphoramidites (**S3**=Q; **S6**=Z; **S8**=Y) we successfully synthesized six microRNA pools (P1-P6) using automated tandem oligonucleotide synthesis. As shown in Table 5.1, P1 consists of miR129-1-5p and miR31; P2 consists of miR129-15p, miR31 and miR206 while; P3 has

all four miRNAs (miR129-15p, miR31, miR206 and miR27b-3p). P4-P6 are similar to P1-P3 differentiating by having the branched linker S3 in between each cleavable site (S8 and S6).

Pool	5'→3' (sequence synthesized)	Calculated mass (g/mol)	Found mass (g/mol)	Purity (% by miRNA)	Estimated Yields
P1	[miR129-5p]-(ZY)-[miR31]	6108.7, 6859.3	6109.0, 6858.9	27.6%(129-5p), 38.3%(31)	24.7%(129-5p), 24.2% (31)
P2	[miR129-5p]-(ZY)-miR31-(ZY)-miR206	6108.7, 6859.3, 7286.5	6108.3, 6859.9, 7284.6	20%(129-5p), 44.7%(31/206)	18% (129-5p) -
P3	[miR129-5p]-(ZY)-[miR31]-(ZY)-[miR206]-(ZY)-[miR27b-3p]	6108.7, 6859.3, 7286.5, 6428	6107.7, 6860.9, 7289.2, 6428.7	16.3%(129-5p), 37.5%(31/206), 20.7%(27b-3p)	26.9%(129-5p), - 26.2% (27b)
P4	[miR129-5p]-(ZQY)-[miR31]	6108.7, 6859.3	6109.4, 6861.3	32%(129-5p), 63.4% (31)	6.1%(129-5p), 22.2% (31)
P5	[miR129-5p]-(ZQY)-[miR31]-(ZY)-[miR206]	6108.7, 6859.3, 7286.5	6110.3, 6859.2, 7283.9	28.5%(129-5p), 48.9% (31/206)	8.0 % (129-5p), -
P6	[miR129-5p]-(ZQY)-[miR31]-(ZY)-[miR206]-(ZQY)-[miR27b-3p]	6108.7, 6859.3, 7286.5, 6428	6109.3, 6865.0, 7288.9, 6426.9	19.8%(129-5p), 28.2%(31/206), 25.0%(27b-3p)	1.9% (129-5p), - 17.6% (27b)

Table 5.1 Characterization of microRNA pools. List of synthesized pools and miRNA in pools with expected and observed masses, purity and estimated yields after purification based on calibration curve from controls (Table S5.3) and peak purity. S3=Q; S6=Z; S8=Y.

Trityl release values were followed during the solid-phase synthesis (**Figure S5.20**). According to this, the coupling efficiencies of phosphoramidites **S6**, **S8** and **S3** were all nearly 100%. However, incorporation of branched linker did not double the trityl release values as expected. This could indicate that not both hydroxyl groups were coupled completely, in spite of increased coupling time and excess of phosphoramidite.

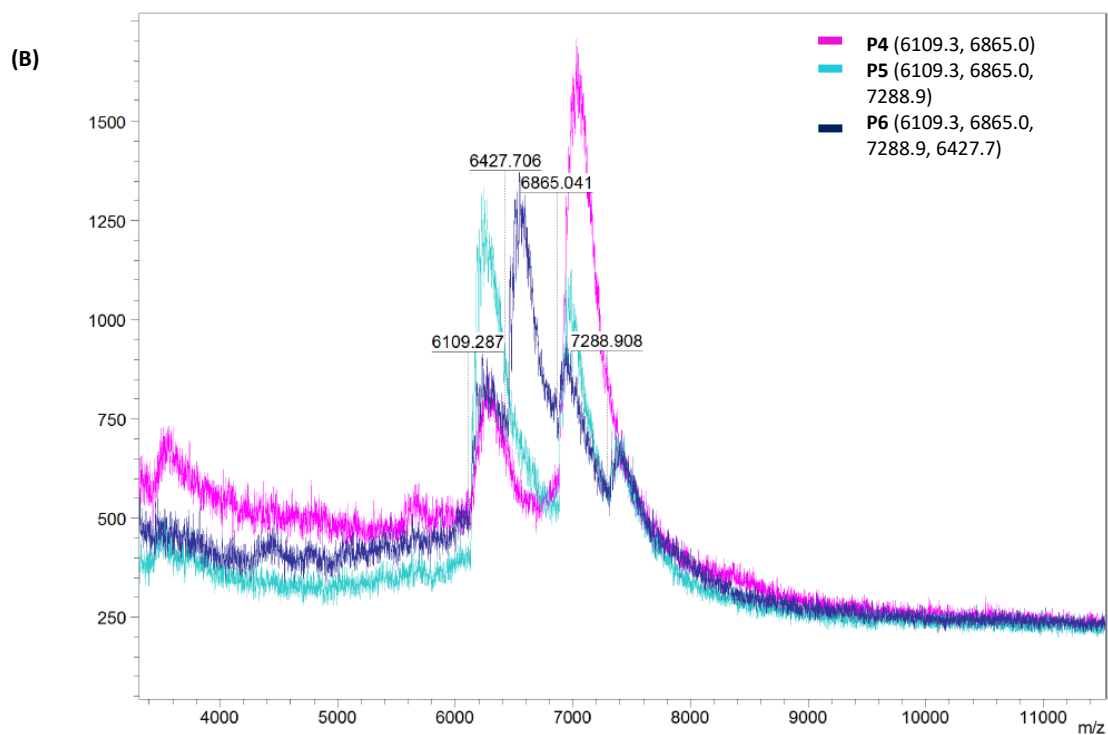
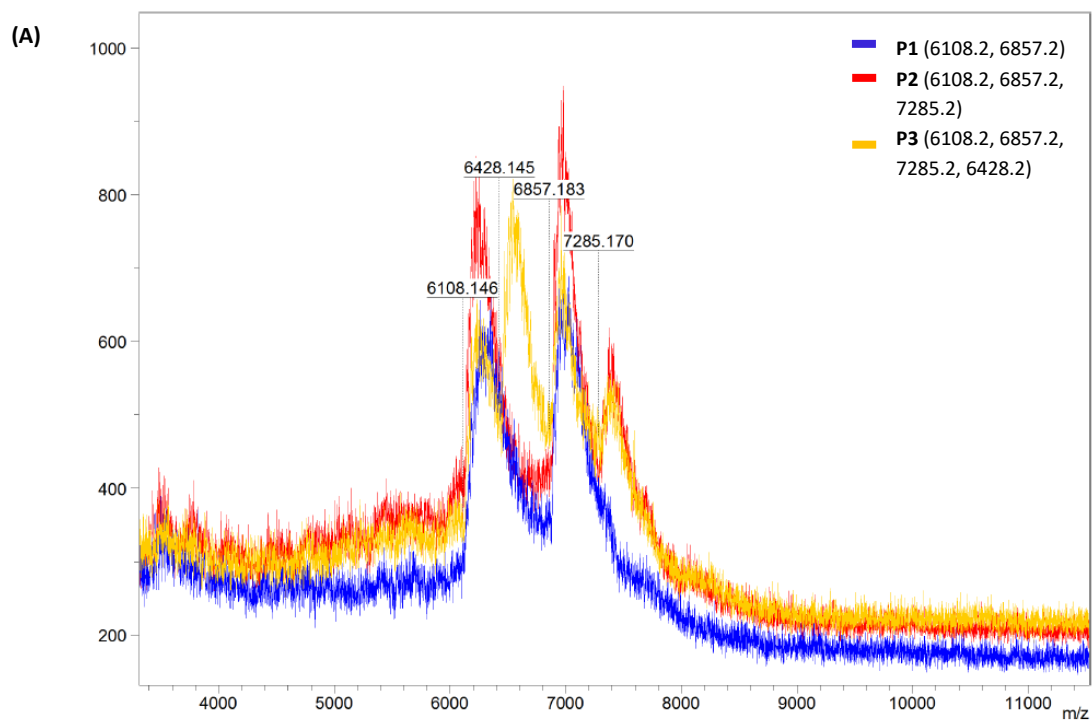
5.2.4 Pool characterization

Characterization of P1-P6 was done using ultra high pressure liquid chromatography (UHPLC) and matrix associated laser desorption ionization with time of flight (MALDI-TOF). Identification of each miRNA in each pool was confirmed from MALDI-TOF spectra of the crude pools. However, the UHPLC trace from crude showed many failed sequences. Therefore, it was decided to isolate individual microRNA from the pools and characterize them independently. miR206 and miR31 had the exact same retention time on UHPLC (**Figure 5.5**). The oligonucleotides were

purified as a pool. Furthermore, we made standard curves of individual microRNA controls (**Figure S5.22**) to estimate yield of each miRNA in the pools (**Table 5.1**). Due to the retention time, we calculated miR206 and miR31 as one peak in P2, P3, P5 and P6. As seen in **Table 5.1** each miRNA in the linear pools (P1-P3) all result in 25% yield after purification.

Notably, miR129-1-5p is the last synthesized ON in all pools which has the same yield as the first one synthesized in P3 (miR27-3p), even though it has the longest linear consecutive sequence. Therefore, we conclude that the 2'/3'-OAc nucleotide phosphoramidites (**S6 and S8**) can be used successfully in combination to make miRNA pools.

By incorporating a branched linker in each cleavable site (P4-P6) we aimed to increase the yield of the sequences. However, from the calculated yields in **Table 5.1** we see that it has the opposite effect compared to the linear pools. The yields of miR129-1-5p in P4-P6 are 6.1, 8 and 1.9%, which theoretically should have been twice, four and eight times higher than in linear pool. Even though we are generating more sites to react, we believe that branching blocks reaction sites leading to failure sequences. Therefore, we confirm that the linear strategy has beneficial outcome compared to “dendrimer” tandem synthesis approach.



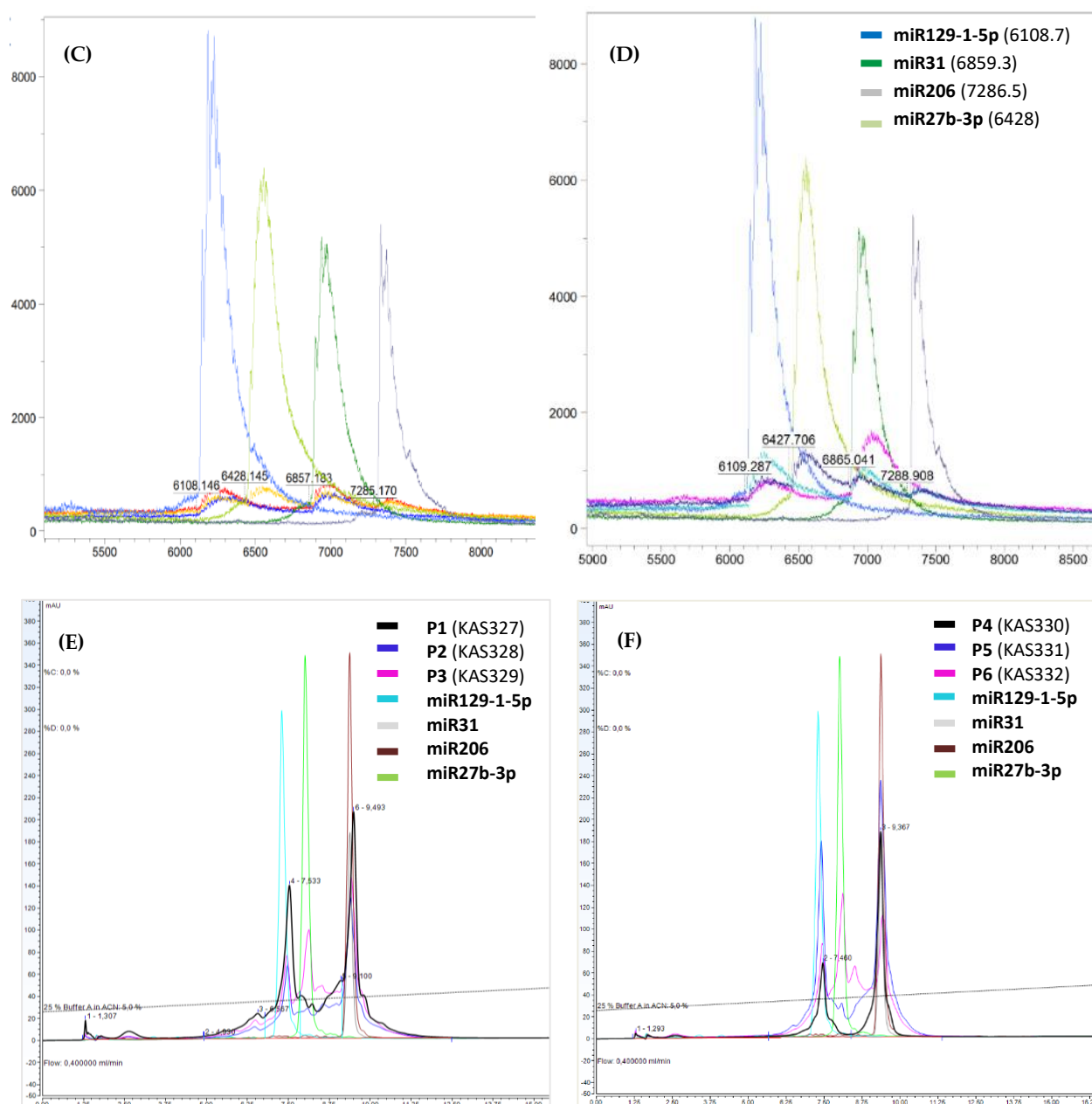


Figure 5.5: Characterization of miRNA pools (P1-P6) compared with controls; MALDI-TOF spectra of (A) P1-P3; (B) P4-P6; (C) P1-P3 with control miRNA; (D) P4-P6 with control miRNA; (E) HPLC trace of P1-P6 with control miRNA; HPLC trace P4-P6 with controls.

5.3 Conclusion

In this work, we report making pools of miRNA by integration of synthesized cleavable 2'/3'OAc nucleotide phosphoramidites using tandem oligonucleotide solid-phase synthesis. We synthesized two 2'/3'OAc-phosphoramidite reagents which combined gives a cleavable site for microRNA pool synthesis. Two, three and four consecutive microRNAs (miR129-1-5p, miR31, miR206, miR27b-3p) were successfully synthesized using linear pool strategy. The identity of microRNA products was verified using UHPLC and MALDI-TOF.

Furthermore, we investigated the effect of a branched molecule **S3** between **S8** and **S6** to generate dendritic structures, aiming for increased yield of the last synthesized ON (ON_N). However, even with sufficient coupling of **S3**, the overall yields were not improved compared to the linear approach.

Incorporation of pools or parallel oligonucleotide synthesis is still not an applied field in big scale oligonucleotide synthesis on solid-phase support. Nevertheless, multiple RNA therapeutic candidates consist of two or more modified oligomers in size of 20mers such as siRNA, miRNA, and heteroduplex antisense oligonucleotides. Therefore, we believe there is a need for tandem synthesis technology. Our cleavable cite approach is a promising way to make microRNA pools and can contribute to the synthesis of RNA therapeutics in the future.

5.4 Experimentals

5.4.1 General experimental methods

All commercial starting materials and solvents were used without further purification. Anhydrous solvents were obtained from a PureSolv™ MD-7 Solvent Purification System, Innovative Technology. TLC was performed using Merck aluminum sheets covered with silica (C-60 F₂₅₄) and visualized at 254 nm or stained with KMnO₄ or H₂SO₄ stains. Purification by flash chromatography was carried out using silica gel 60 (0.040 – 0.063 mm, Merck). ¹H NMR, ¹³C NMR and ³¹P NMR spectra were recorded on Bruker Ascend 400 with a Prodigy cryoprobe (operating at 400 MHz for proton, 101 MHz for carbon and 162 MHz for phosphor) at 300 K. Chemical shifts are reported in ppm (δ) downfield from TMS (δ=0). Spectra were calibrated relative to residual solvent peak: ¹H NMR (CDCl₃), 7.26 ppm, ¹³C NMR (CDCl₃), 77.16 ppm; ¹H NMR (DMSO-*d*₆), 2.50 ppm, ¹³C NMR (DMSO-*d*₆), 39.52 ppm; ¹H NMR (CD₃OD), 3.31 ppm, ¹³C NMR (CD₃OD), 49.00 ppm; ¹H NMR (acetone-*d*₆), 2.05 ppm, ¹³C NMR (acetone-*d*₆), 29.84, 206.26 ppm. Coupling constants (*J*) are given in hertz, and notation represents the multiplicities singlet (s), doublet (d), doublet of doublets (dd), triplet (t), quartet (q), multiplet (m) and broad signals (bs). Assignment was based on ¹H-NMR, ¹³C-NMR, ³¹P-NMR, ¹H-¹H COSY and HSQC NMR spectra.

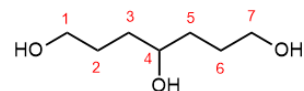
UHPLC-HRMS was performed on an Agilent Infinity 1290 UHPLC system (Agilent Technologies, Santa Clara, CA, USA) equipped with a diode array detector. MS detection on Agilent 6545 QTOF MS equipped with Agilent Dual Jet Stream electrospray ion source.

Liquid chromatography mass spectra (UPLC-MS) were recorded on a Waters AQUITY UPLC system equipped with PDA and SQD electrospray MS detector. Column: Thermo accucore C18 2.6 μm, 2.1 x 50 mm. Column temperature: 50°C and Flowrate: 0.6 mL/min. Solvent A: 0.1% formic acid in water, solvent B: 0.1% formic acid in acetonitrile OR solvent A: 10nM NH₄Ac in water, solvent B: 10mM NH₄Ac 90:10 acetonitrile:water. Gradient short run (2.6 min) or long run (5 min): 5-100% B in A.

5.4.2 Synthesis of monomers

5.4.2.1 Heptane-1,4,7-triol (S1)

S1 was prepared according to procedure¹⁷⁰. To a solution of diethyl 4-oxoheptanedioate (6.1 g, 26.49 mmol) in THF (150 mL) was added

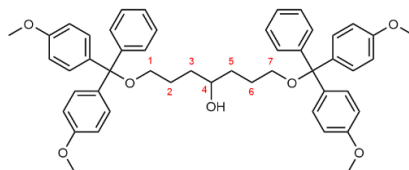


LiAlH₄ 2 M solution in THF (4.55 g, 120 mmol) portion wise at 0 °C. The reaction mixture was stirred overnight upon which H₂O (4.55 mL) was added dropwise to the solution at 0 °C over 20 min, followed by dropwise addition of 15 % NaOH (4.55 mL) at 0 °C over 20 min, and finally H₂O

(13.65 mL) to quench the reaction. The resulting suspension was filtered under suction and concentrated under reduced pressure to give a crude product. The filtrate was purified by column chromatography (20% MeOH:DCM) and concentrated by reduced pressure into the title compound **S1** as a colorless oil (2.86 g, 73 %). $R_f = 0.42$ (MeOH:DCM, 20:80); $^1\text{H NMR}$ (400 MHz, Methanol- d_4) δ 3.62 – 3.51 (m, 5H, H1, H4, H7), 1.76 – 1.36 (m, 8H, H2, H3, H5, H6); $^{13}\text{C NMR}$ (101 MHz, Methanol- d_4) δ 70.7 (C4), 61.7 (C1,C7), 33.4 (C2,C6), 28.5 (C3,C5); **HRMS** (ESI) calc. for $\text{C}_7\text{H}_{16}\text{O}_3$ $[\text{M} + \text{Na}]^+$ 171.0997, found 171.0992 m/z.

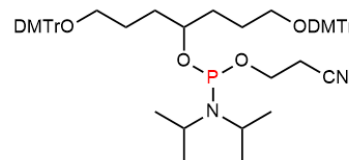
5.4.2.2 1,7-di-*O*-(4,4'-dimethoxytrityl)heptane-1,4,7-triol (**S2**)

S2 was prepared according to the procedure¹⁷⁰. To a solution of heptane-1,4,7-triol (1.10 g, 7.42 mmol) in anhydrous pyridine (30 mL) was added DMT chloride (10.56 g, 31.16 mmol). The reaction was stirred overnight at rt., upon which was partitioned between brine (150 mL) and DCM (4 x 150 mL). The organic phases were collected, reduced under pressure and purified by column chromatography (10-35% EtOAc in heptane) to obtain the title compound **S2** as a white foam (5.08 g, 91 %). $R_f = 0.21$ (EtOAc/Hept, 30:70); $^1\text{H NMR}$ (400 MHz, Acetone- d_6) δ 7.53 – 7.15 (m, 18H, 2xDMTr), 6.90 – 6.81 (m, 8H, 2xDMTr), 3.77 (s, 12H, OCH₃), 3.54 – 3.43 (m, 1H, H4), 3.07 (t, $J = 6.5$ Hz, 4H, H1,H7), 1.86 – 1.36 (m, 8H, H2,H3,H5,H6); $^{13}\text{C NMR}$ (101 MHz, Acetone- d_6) δ 159.5, 146.7, 137.5, 130.9, 129.0, 128.5, 127.4, 113.8 (DMTr), 86.5 (DMTr-C), 64.2 (C1,C7), 55.5 (OCH₃), 35.2 (C3,C5), 27.1 (C2,C6); **HRMS** (ESI) calc. for $\text{C}_{49}\text{H}_{52}\text{O}_7$ $[\text{M} + \text{Na}]^+$ 775.3611, found 775.3605 m/z.



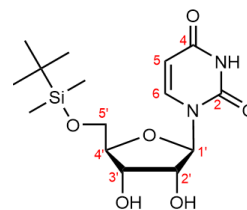
5.4.2.3 1,7-di-*O*-dimethoxytrityl heptane-4-*O*-(2-cyanoethyl)-*N,N*-diisopropylphosphoramidite (**S3**)

S3 was prepared according to the procedure¹⁷⁰. Compound **S3** was prepared by co-evaporating **S2** (400 mg, 531.25 μmol) with anhydrous DCE (7 mL) and redissolved in anhydrous DCM (4 mL). 2-Cyanoethyl *N,N,N',N'*-tetraisopropylphosphoramidite (320 mg, 340 μL , 1.06 mmol) and diisopropylammonium tetrazolide (182 mg, 1.06 mmol) were added, and the reaction was stirred under argon at rt overnight. The mixture was concentrated under reduced pressure and loaded onto silica column. The crude product was eluted with 10-35% EtOAc in heptane and reduced under pressure to obtain the title compound **S3** as a white foam (485.0 mg, 95%). $R_f = 0.36$ (EtOAc/Hept, 30:70); $^{31}\text{P NMR}$ (162 MHz, DMSO- d_6) δ 146.67; **HRMS** (ESI) calc. for $\text{C}_{58}\text{H}_{69}\text{N}_2\text{O}_8\text{P}$ $[\text{M} + \text{Na}]^+$ 975.4689, found 975.4684 m/z.

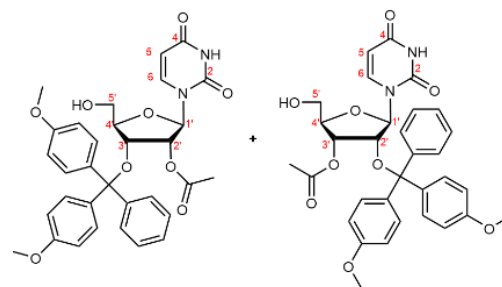


5.4.2.4 5'-*O*-*tert*-Butyldimethylsilyluridine (S4)

S4 was prepared according to the procedure¹⁸¹. Compound **S4** was prepared from uridine (3.02 g, 12.28 mmol) that was co-evaporated with anhydrous pyridine (20 mL) and dissolved in anhydrous DMF (40 mL). Imidazole (1.87 g, 27.27 mmol) was added, and the reaction mixture stirred until all solids had dissolved. *tert*-Butyldimethylsilyl chloride (2.07 g, 13.64 mmol) was added, and the solution was stirred for 3.5 h. The mixture was then poured into water (200 L) and stirred vigorously for 30 min. The white solid that formed was collected by vacuum filtration through a sintered glass funnel to give the title compound **S4** as a white solid (3.69 g, 84%). R_f = 0.23 (5% MeOH in DCM); $^1\text{H NMR}$ (400 MHz, DMSO- d_6) δ 11.35 (s, 1H, NH), 7.79 (d, J = 8.1 Hz, 1H, H6), 5.77 (d, J = 4.9 Hz, 1H, H1'), 5.58 (dd, J = 8.1, 2.1 Hz, 1H, H5), 5.47 (d, J = 5.7 Hz, 1H, OH), 5.12 (d, J = 5.3 Hz, 1H, OH), 4.05 – 3.87 (m, 3H, H2', H3', H4'), 3.84 (dd, J = 11.6, 2.9 Hz, 1H, H5'), 3.73 (dd, J = 11.4, 3.1 Hz, 1H, H5''), 0.90 (s, 9H, 3xCH₃, TBDMS), 0.09 (s, 6H, 2xCH₃, TBDMS); $^{13}\text{C NMR}$ (101 MHz, DMSO- d_6) δ 163.1 (C4), 150.6 (C2), 140.3 (C6), 101.6 (C5), 88.0 (C1'), 84.1 (C4'), 73.8 (C2'), 69.6 (C3'), 62.7 (C5'), 25.9 (3xCH₃, TBDMS), 18.1 (TBDMS-C), -5.5 (2xCH₃, TBDMS); LCMS (ESI) calc. for C₁₅H₂₆N₂O₆Si [M+H]⁺ 359.1633 found 359.6 m/z.

5.4.2.5 2'-*O*-Acetyl-3'-*O*-(4,4'-dimethoxytrityl)uridine and 3'-*O*-Acetyl-2'-*O*-(4,4'-dimethoxytrityl)uridine (S5a+ S5b)

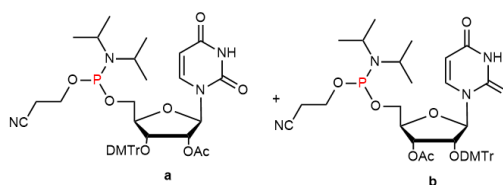
S5a+S5b was prepared according to the procedure¹⁸². 5'-*O*-*tert*-butyldimethylsilyluridine **S4** (13.0 g, 35 mmol) was co-evaporated with anhydrous pyridine (100 mL) and then redissolved in anhydrous pyridine (100 mL).



DMT-chloride (12.0 g, 36 mmol) was added, and the solution was stirred at rt for 48 h. Ac₂O (20 mL) and *N*-methylimidazole (2.0 mL, 25 mmol) were added, and the solution was stirred at rt for an additional 24 h, then concentrated under reduced pressure. The resultant residue was dissolved in EtOAc (300 mL), washed with 1 N citric acid (300 mL), brine (200 mL), dried over MgSO₄, and concentrated under reduced pressure. The resulting oil was dissolved in THF (100 mL), treated with 1 M TBAF in THF (50 mL), and stirred at rt. for 4 h. The solvent was then removed by rotary evaporation, and the residue dissolved in EtOAc (200 mL), washed with saturated aqueous NaHCO₃ (200 mL), brine (200 mL), dried over MgSO₄, and concentrated. The crude product was purified by column chromatography (0-5% MeOH in DCM) to give the title compound **S5** as a white foam (11.9 g; 56%) and regioisomeric mixture in ratio 80:20 (S5a:S5b) from NMR. R_f = 0.13 (EtOAc/Hept, 60:40); $^1\text{H NMR}$ (400 MHz, DMSO- d_6) δ 11.50 (s, 1H, NH, 5a+5b), 7.78 (d, J = 8.2 Hz, 0.2H, H6, 5b), 7.57 (d, J = 8.2 Hz, 0.8H, H6, 5a), 7.50 – 7.42 (m, 0.4H,

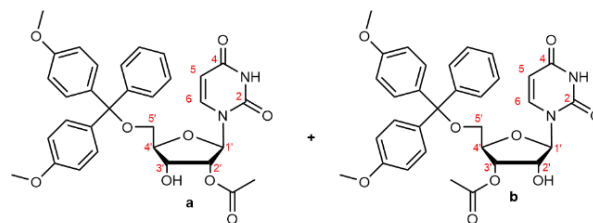
aromatic), 7.36 – 7.30 (m, 3H, aromatic), 7.25 – 7.17 (m, 6H, aromatic), 6.97 – 6.89 (m, 1H, aromatic), 6.86 – 6.75 (m, 3H, aromatic), 6.27 (d, $J = 8.0$ Hz, 0.8H, H1', 5a), 6.17 (d, $J = 7.0$ Hz, 0.2H, H1', 5b), 5.68 (d, $J = 8.1$ Hz, 0.2H, H5, 5b), 5.56 (d, $J = 8.1$ Hz, 0.8H, H5, 5a), 5.22 (t, $J = 4.3$ Hz, 0.8H, OH, 5a), 5.15 (t, $J = 4.9$ Hz, 0.2H, OH, 5b), 4.99 (dd, $J = 7.0, 5.4$ Hz, 0.2H, H2', 5b), 4.30 (dd, $J = 5.4, 2.1$ Hz, 0.2H, H3', 5b), 4.25 (dd, $J = 8.1, 5.2$ Hz, 0.8H, H2', 5a), 3.86 (d, $J = 1.9$ Hz, 0.8H, H4', 5a), 3.72 (2xs, 6H, 2xOCH₃, 5a+5b), 3.64 (d, $J = 5.2$ Hz, 0.8H, H3', 5a), 3.36 – 2.92 (m, 2.2H, H5'/H5'', 5a+5b, H4', 5b), 2.13 (s, 0.6H, acetyl, 5b), 2.04 (s, 2.4H, acetyl 5a); ¹³C NMR (101 MHz, DMSO-*d*₆) δ 169.7 (C=O, 5a+5b), 163.0 (C4, 5a+5b), 158.5, 158.4 (aromatic, 5a+5b), 151.0 (C2, 5a+5b), 144.6 (aromatic), 140.4, 139.7 (C2, 5a+5b), 135.6, 135.0, 134.9, 130.2, 130.1, 129.8, 129.7, 128.0, 127.9, 127.6, 127.1, 113.4, 113.2 (aromatic, 5a+5b), 102.9 (C5, 5a), 102.5 (C5, 5b), 86.4 (C1', 5b), 86.3 (C1', 5a), 85.4 (C4', 5a), 84.7 (C4', 5b), 74.0 (C2', 5a), 73.4 (C2', 5b), 72.8 (C3', 5a), 72.6 (C3', 5b), 61.2 (C5', 5a+5b), 55.1, 55.1, (OCH₃, 5a+5b), 20.9 (acetyl, 5a), 20.7 (acetyl, 5b); HRMS (ESI) calc. for C₃₂H₃₂N₂O₉ [M + Na]⁺ 611.2006, found 611.2000.

5.4.2.6 5'-O-(N,N-Diisopropyl-2-cyanoethylphosphoramidite)-2'-O-acetyl-3'-O-(4,4'-dimethoxytrityl) uridine and 5'-O-(N,N-Diisopropyl-2-cyanoethylphosphoramidite)-3'-O-acetyl-2'-O-(4,4'-dimethoxytrityl) uridine (S6a+S6b)



S6 was prepared according to the procedure¹⁸². 2'-O-acetyl-3'-O-(4,4'-dimethoxytrityl)-5'-O-uridine (S5a+S5b) (400 mg, 0.68 mmol) was co-evaporated with anhydrous DCE (5 mL) and redissolved in DCM (5 mL). 2-Cyanoethyl *N,N,N',N'*-tetraisopropylphosphoramidite (410 mg, 430 μ l, 1.36 mmol) and diisopropylammonium tetrazolide (237 mg, 1.36 mmol) were added and the reaction mixture was stirred under argon at rt overnight. The mixture was concentrated in vacuo and purified by flash chromatography (30-60% EtOAc/Hept) to give the title compound S6a+S6b as a white foam (422 mg, yield 78%). $R_f = 0.45, 0.41, 0.34, 0.30$ in 60% EtOAc/Hept; ³¹P NMR (162 MHz, Chloroform-*d*) δ 149.46, 148.86, 146.87, 145.66; HRMS (ESI) calc. for C₄₁H₄₉N₄O₁₀P [M + H]⁺ 789.3259, found 789.3259.

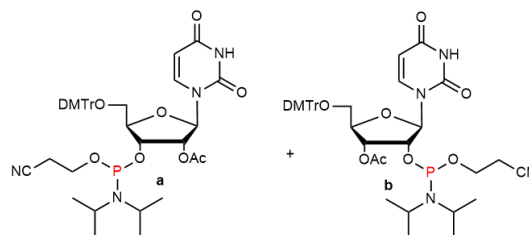
5.4.2.7 2'/3'-O-acetyl-5'-O-(4,4'-dimethoxytrityl)uridine (S7a+S7b)



S7 was prepared according to procedure¹⁷¹. To a solution of commercially available 5'-O-(4,4'-dimethoxytrityl)uridine (1.5 g, 2.74 mmol) in dry THF (10 mL) was added anhydrous pyridine (222 μ l, 2.74 mmol) followed by acetyl chloride (195 μ l, 2.74 mmol) at 0 °C under argon. The mixture was warmed to rt., stirred for 3 h and quenched with saturated NaHCO₃ (15 mL). The organics were extracted with DCM (5x20 mL), dried over

MgSO₄ and concentrated under reduced pressure. The residue was purified using column chromatography (20-100% EtOAc in Heptane), co-evaporated with dry DCM to give a regioisomeric mixture of the title compound **S7a** and **S7b** as a white foam (429 mg, 26.6%). The isomers were not isolated but found in ratio 0.4:0.6 (7a:7b) in the UV-trace from LC-MS and confirmed by integration of H1' in NMR spectra. *R_f* = 0.39 (EtOAc/Hept, 80:20); ¹H NMR (400 MHz, DMSO-*d*₆) δ 11.43 (s, 1H, NH), 7.70 (2xd, *J* = 8.1 Hz, 1H, H6,7a+7b), 7.44 – 7.20 (m, 9H, aromatic, 7a+ 7b), 6.95 – 6.87 (m, 4H, aromatic, 7a+7b), 5.92 (d, *J* = 4.5 Hz, 0.4H, H1', 7a), 5.82 (d, *J* = 5.8 Hz, 0.6H, 2'-OH, 7b), 5.76 (d, *J* = 5.6 Hz, 0.6H, H1', 7b), 5.58 (d, *J* = 6.2 Hz, 0.4H, 3'-OH, 7a), 5.46 (d, *J* = 7.9 Hz, 0.6H, H5, 7b), 5.41 (d, *J* = 8.0 Hz, 0.4H, H5, 7a), 5.27 (t, *J* = 5.1 Hz, 0.4H, H2',7a), 5.14 (t, *J* = 5.1 Hz, 0.6H, H3', 7b), 4.45 – 4.19 (m, 1H, H3' 7a, H2' 7b), 4.12 (q, *J* = 4.2 Hz, 0.6H, H4', 7b), 4.01 – 3.97 (m, 0.4H, H4', 7a), 3.75 (s, 6H, 2xOCH₃, 7a+7b), 3.32 – 3.18 (m, 2H, H5'/H5'', 7a+7b), 2.09 (s, 1.2H, CH₃, 7a), 2.07 (s, 1.8H, CH₃, 7b); ¹³C NMR (101 MHz, DMSO-*d*₆) δ 170.1, 170.0 (acetyl C=O), 163.4, 163.4 (C4), 158.6, 158.6 (aromatic), 150.1 (C2,7a+7b), 145.1, 145.0 (aromatic), 140.5 (C6,7a+7b), 135.7, 135.5, 130.2, 130.2, 128.4, 128.4, 128.2, 128.1, 127.3, 127.3, 113.8, 113.7 (aromatic), 102.9, 102.4 (C5,7b+7a), 88.9 (C1',7b), 86.5 (C1',7a), 86.4 (DMTr-C), 83.4 (C4',7a), 80.7 (C4',7b), 75.6 (C2',7a), 72.4 (C3',7b), 71.7 (C2',7b), 68.7 (C3',7a), 63.5, 63.4 (C5',7a+7b), 55.5 (OCH₃,7a+7b), 21.2, 21.1 (acetyl CH₃,7a+7b); LCMS (ESI) calc. for C₃₂H₃₂N₂O₉ [M-H]⁻ 587.6, found 587.7 m/z for both regioisomers.

5.4.2.8 2'/3'-O-Acetyl-5'-O-(4,4'-dimethoxytrityl)-uridine-3'/2'-O-(2-cyanoethyl-*N,N*-diisopropyl)phosphoramidite (**S8a**+**S8b**)



S8 was prepared according to procedure¹⁷¹.

Compound **S8** was prepared by co-evaporating **S7**

(204 mg, 340 μmol) in a dry round bottom flask with dry DCE (7 ml) and redissolved in dry DCM (5ml) followed by addition of 2-cyanoethyl *N,N,N',N'*-tetraisopropylphosphoramidite (205 mg, 215 μl, 680 μmol) and diisopropylammonium tetrazolide (116 mg, 680 μmol). The reaction was stirred under argon o/n. The reaction mixture was reduced and purified on column using 10-70% (EtOAc in Heptane) as eluent. The pure fractions were collected, reduced and redissolved in dry DCM and finally evaporated to obtain the title compound **S8** as a white foam (220 mg, 82%) and a regioisomeric mixture. *R_f* = 0.45, 0.40 (EtOAc/Hept, 60:40); ³¹P NMR (162 MHz, DMSO-*d*₆) δ 151.78 (**S8b**), 151.16 (**S8b**), 151.04 (**S8a**). Last **S8a**-isomer not collected due to oxidation.

5.4.3 Synthesis of miRNA pools

5.4.3.1 Automated synthesis of microRNA pools

Oligonucleotide pools (P1-P6) were synthesized on Expedite 8909 Nucleic Acid Synthesis System from Applied Biosystems in 0.2 μ mol scale using universal support (CUTAG 1000 Å CPG from Sigma Aldrich). 2'-O-TBDMS and 2'-O-Methyl protected phosphoramidites from Sigma Aldrich (rA(bz), rC(ac), rG(ib), rU) were dissolved in dry MeCN (0.07 M) and put on machine. Reagents for standard automated oligonucleotide synthesis were MeCN, TCA Deblock, DCI Activator 0.25M (4,5-dicyanoimidazole), Oxidizer 0.02M, Cap A (acetic anhydride, THF), Cap B (NMI, pyridine, THF) from Sigma Aldrich. **S3**, **S6** and **S8** were coupled manually by dissolving the respective phosphoramidite (15 mg) with 400 μ l MeCN and 600 μ l activator (DCI) in 1 ml syringe and added to the column in a pace of 50 μ l/min (over 20 min). Trityl values was monitored on the machine. Automated oligonucleotide solid-phase synthesis was run on DMT-off mode on 0.2 μ mol double-coupling program for linear and 1 μ mol double-coupling program for branched with coupling time on 20 min for each cycle.

5.4.3.2 Deprotection, cleavage, quantification and purification of microRNA pools

The column was dried with nitrogen and cleavage from solid support was done using methylamine solution (33 wt. % in absolute ethanol) at 65°C in 2 h, desilylated with triethylamine trihydrofluoride in dry triethylamine and dry *N*-methyl-2-pyrrolidone and precipitated in cold acetone. The microRNA pools were dissolved in RNase free water and identified using an Autoflex speed MALDI-TOF mass spectrometer (Bruker Daltonics, Hamburg, Germany). The oligonucleotides were purified on UHPLC (Dionex Ultimate 3000) using a DNA-Pac RP column (4 μ m) with a 5 - 15 % gradient, 30 min at 75 °C with 0.05 M TEAA in MQ pH 7.4 (buffer A) and acetonitrile (buffer B). Peaks were monitored at 260 nm. Calibration curves were made from control miRNAs purchased from IDT (Table S2.2). Yield estimations and calculations were based on calibration curves and purity peaks from HPLC (**Table S5.3**).

5.5 Supporting Information

5.5.1 NMR-spectra of S1-S8

5.5.1.1 Heptane-1,4,7-triol (S1)

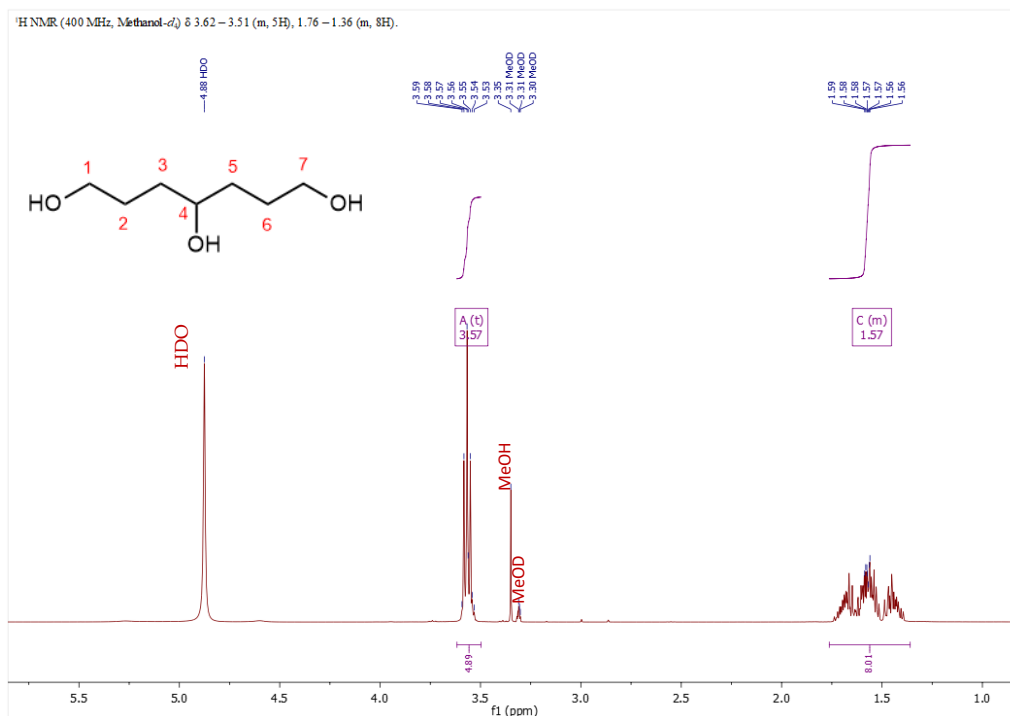


Figure S5.1: ¹H-NMR spectra of Heptane-1,4,7-triol (S1)

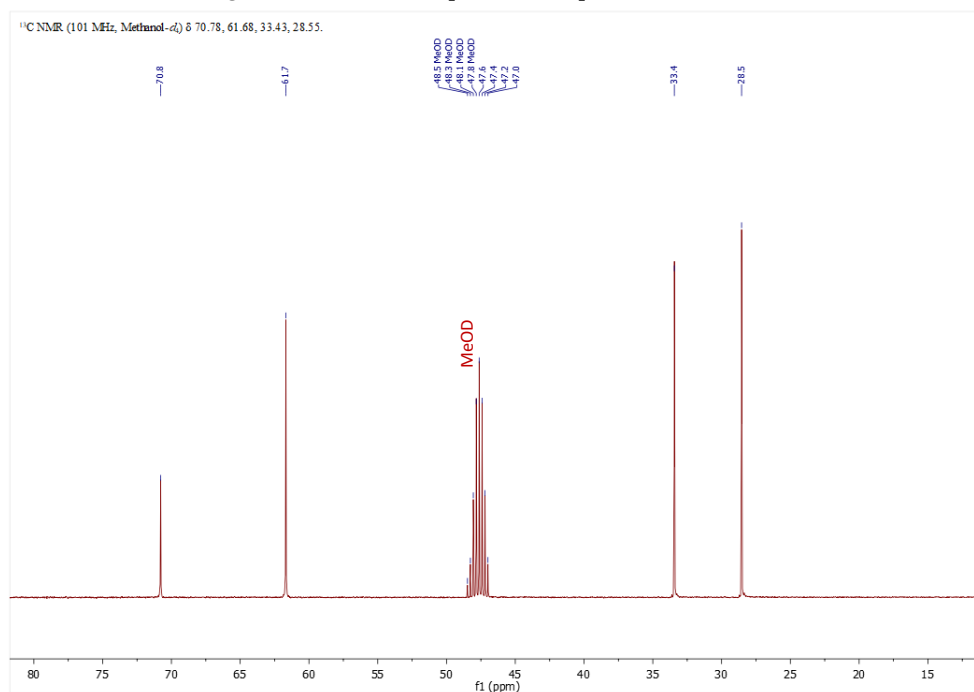
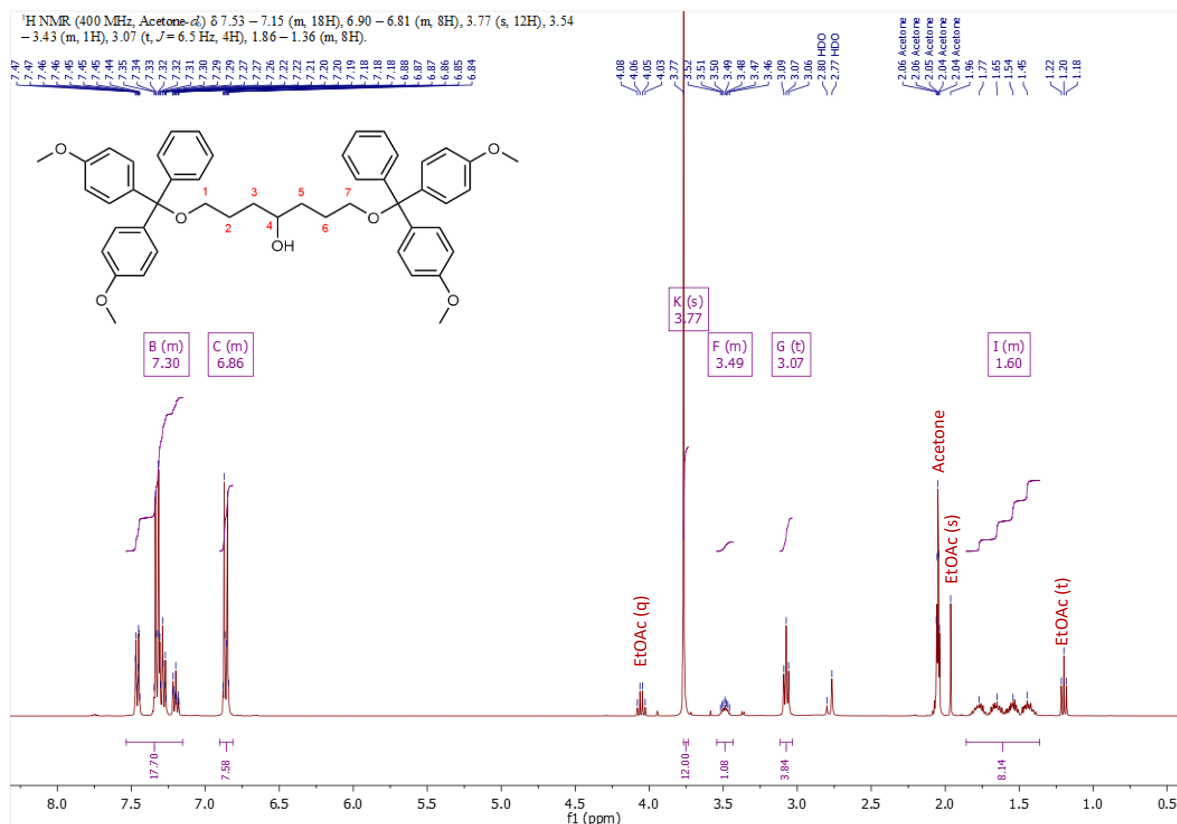
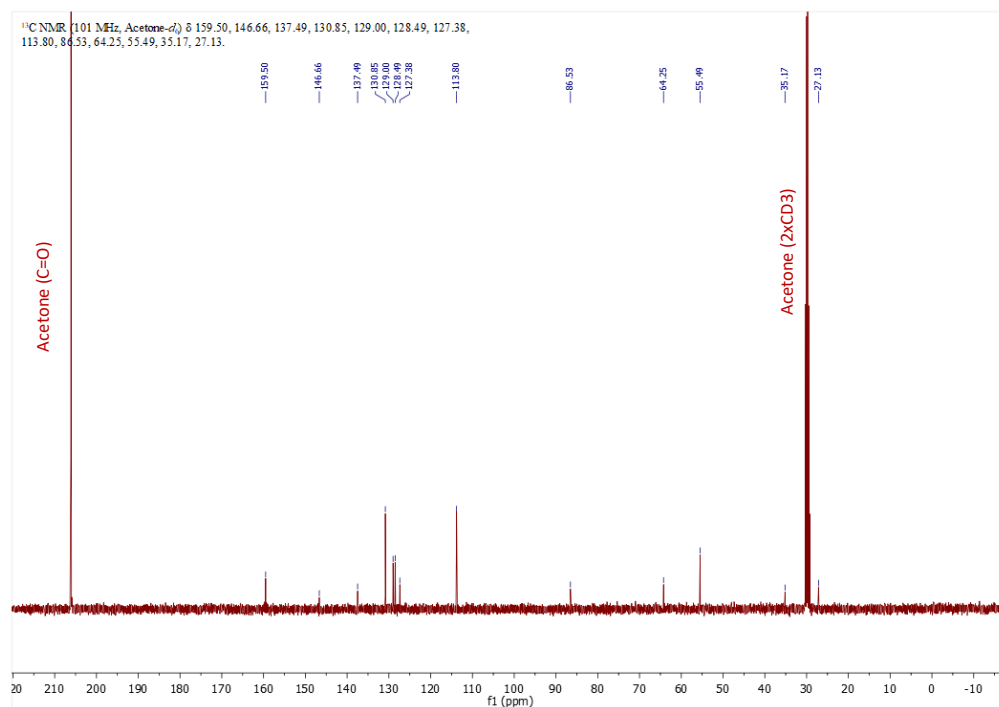


Figure S5.2 ¹³C-NMR spectra of Heptane-1,4,7-triol (S1)

5.5.1.2 1,7-di-*O*-(4,4'-dimethoxytrityl)heptane-1,4,7-triol (S2)Figure S5.3 ¹H-NMR spectra of 1,7-di-*O*-(4,4'-dimethoxytrityl)heptane-1,4,7-triol (S2)Figure S5.4 ¹³C-NMR spectra of 1,7-di-*O*-(4,4'-dimethoxytrityl)heptane-1,4,7-triol (S2)

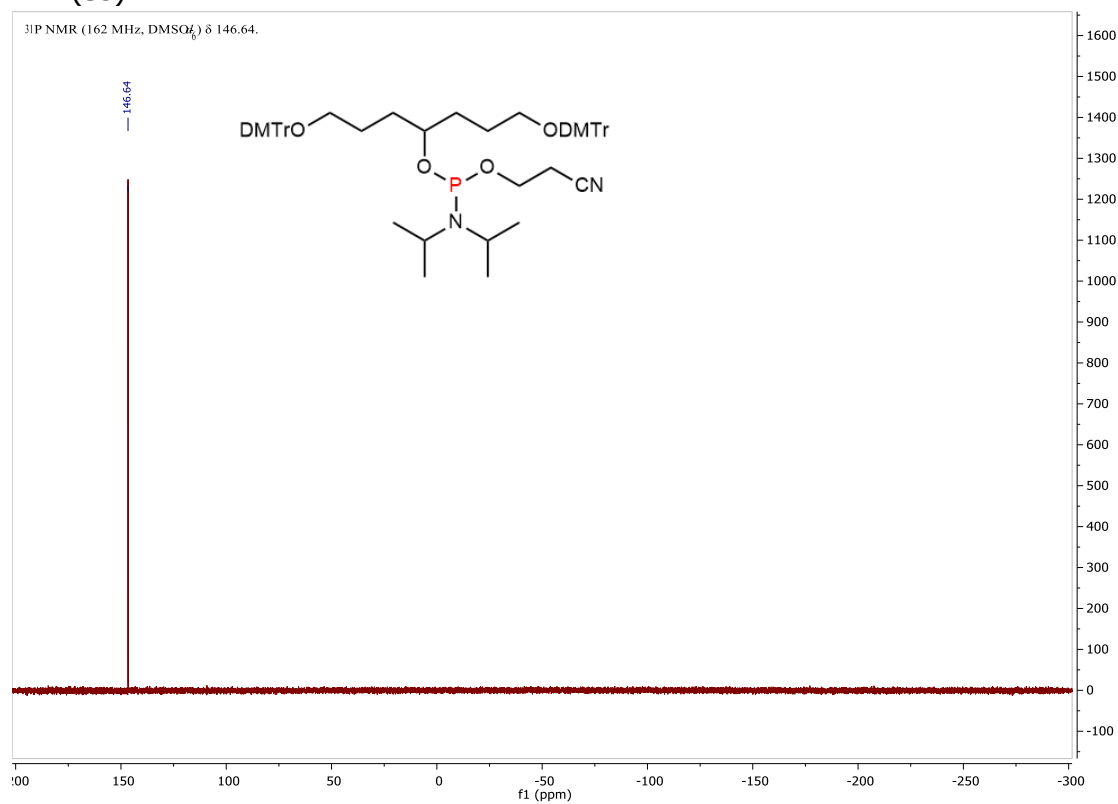
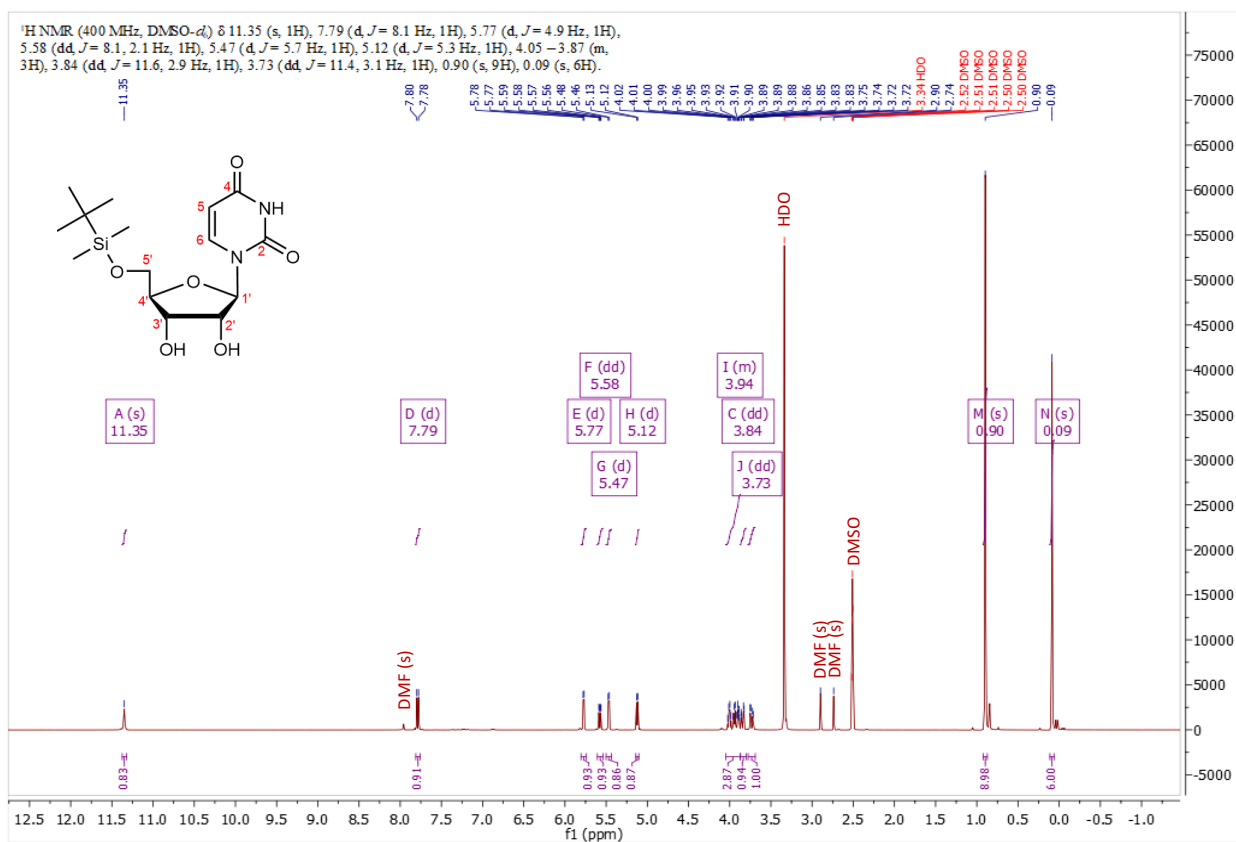
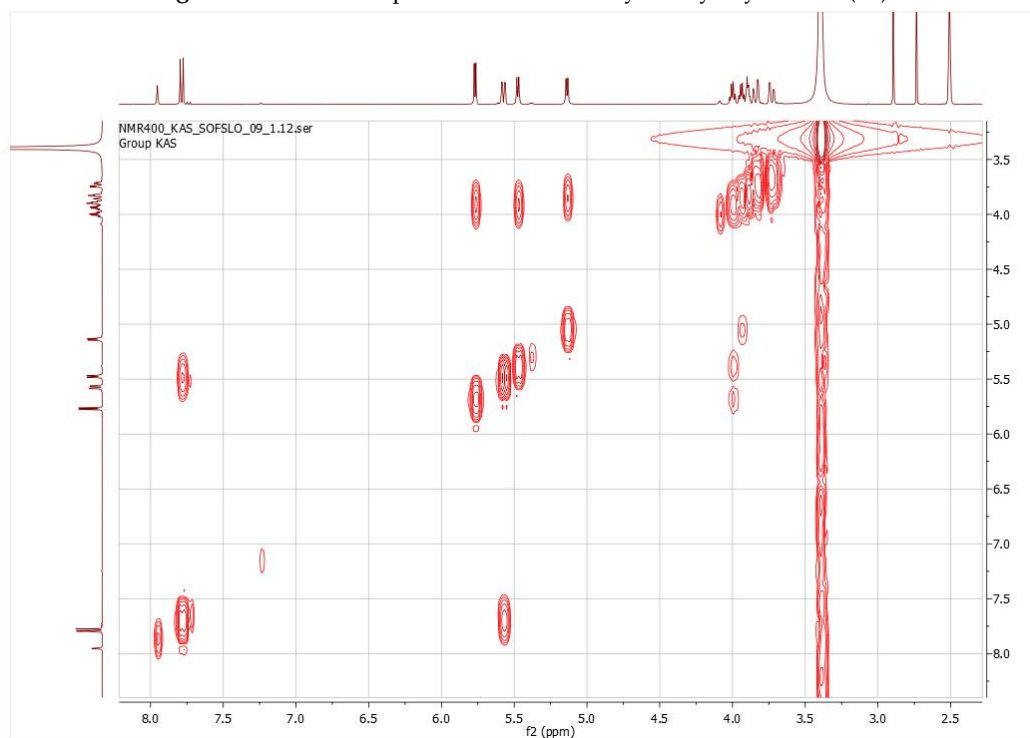
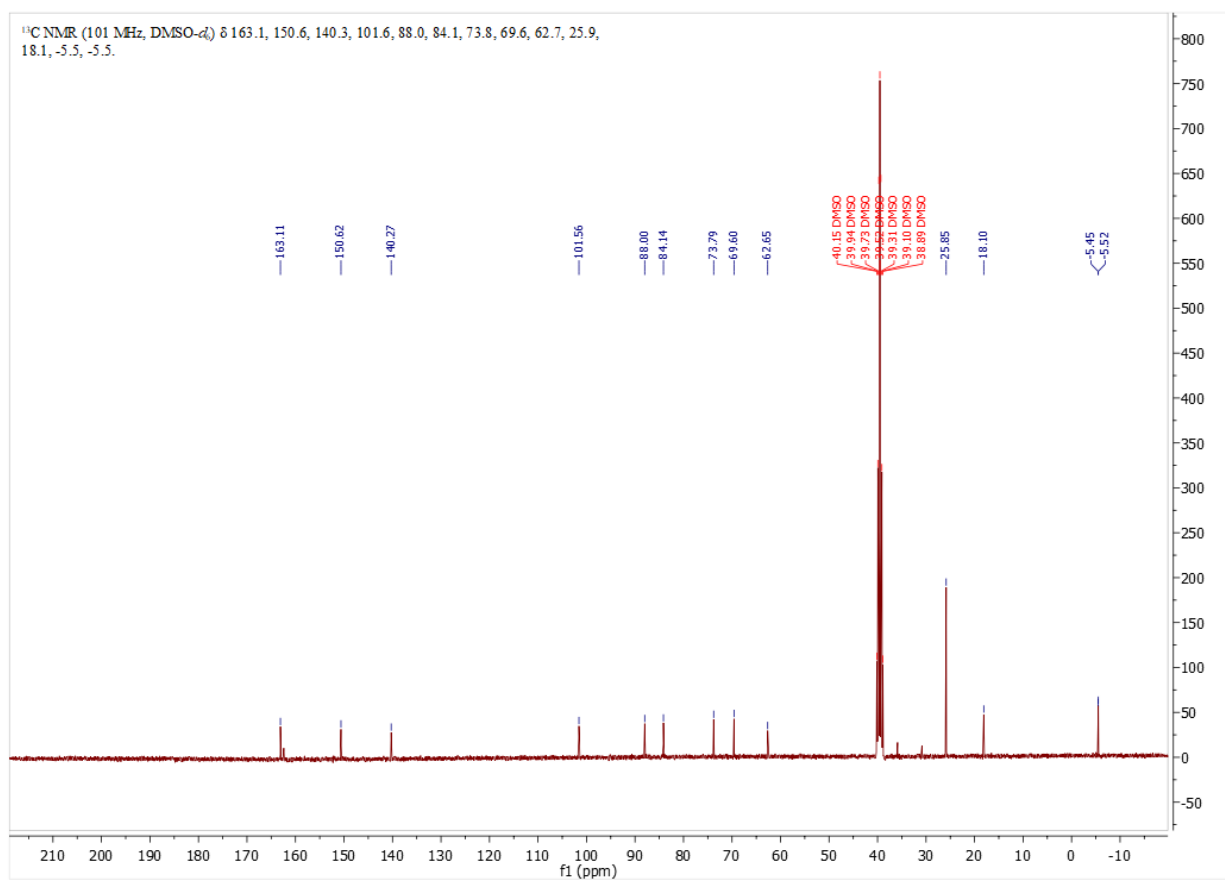
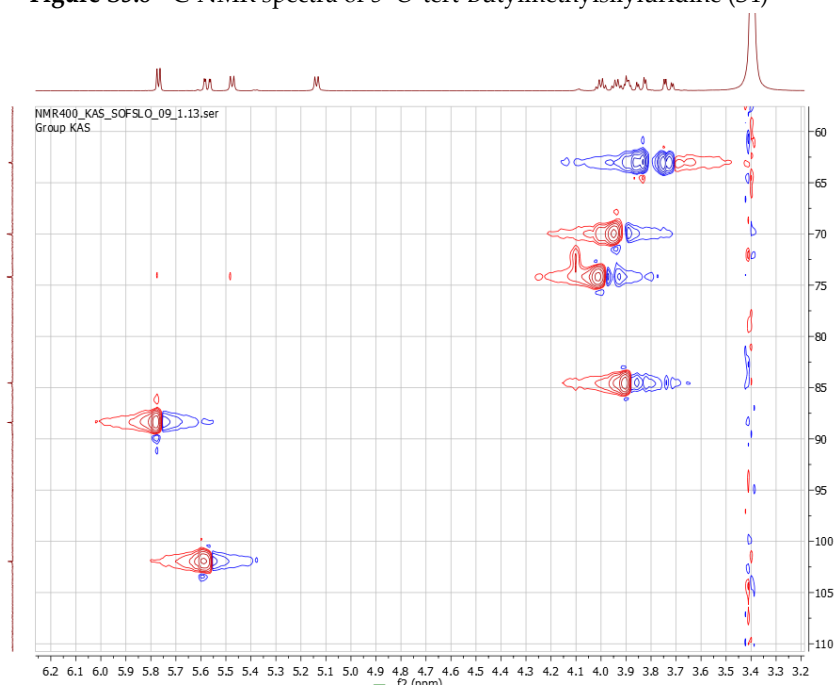
5.5.1.3 1,7-di-*O*-dimethoxytrityl heptane-4-*O*-(2-cyanoethyl)-*N,N*-diisopropylphosphoramidite (S3)

Figure S5.5 ³¹P-NMR spectra of 1,7-di-*O*-dimethoxytrityl heptane-4-*O*-(2-cyanoethyl)-*N,N*-diisopropylphosphoramidite (S3)

5.5.1.4 5'-*O*-tert-Butyldimethylsilyluridine (S4)Figure S5.6 ¹H-NMR spectra of 5'-*O*-tert-Butyldimethylsilyluridine (S4)Figure S5.7 ¹H-¹H COSY-NMR spectra of 5'-*O*-tert-Butyldimethylsilyluridine (S4)

**Figure S5.8** ^{13}C -NMR spectra of 5'-*O*-tert-Butylmethylsilyluridine (S4)**Figure S5.9** HSQC-NMR spectra of 5'-*O*-tert-Butylmethylsilyluridine (S4)

5.5.1.5 2'-O-Acetyl-3'-O-(4,4'-dimethoxytrityl)uridine and 3'-O-Acetyl-2'-O-(4,4'-dimethoxytrityl) uridine (S5a+ S5b)

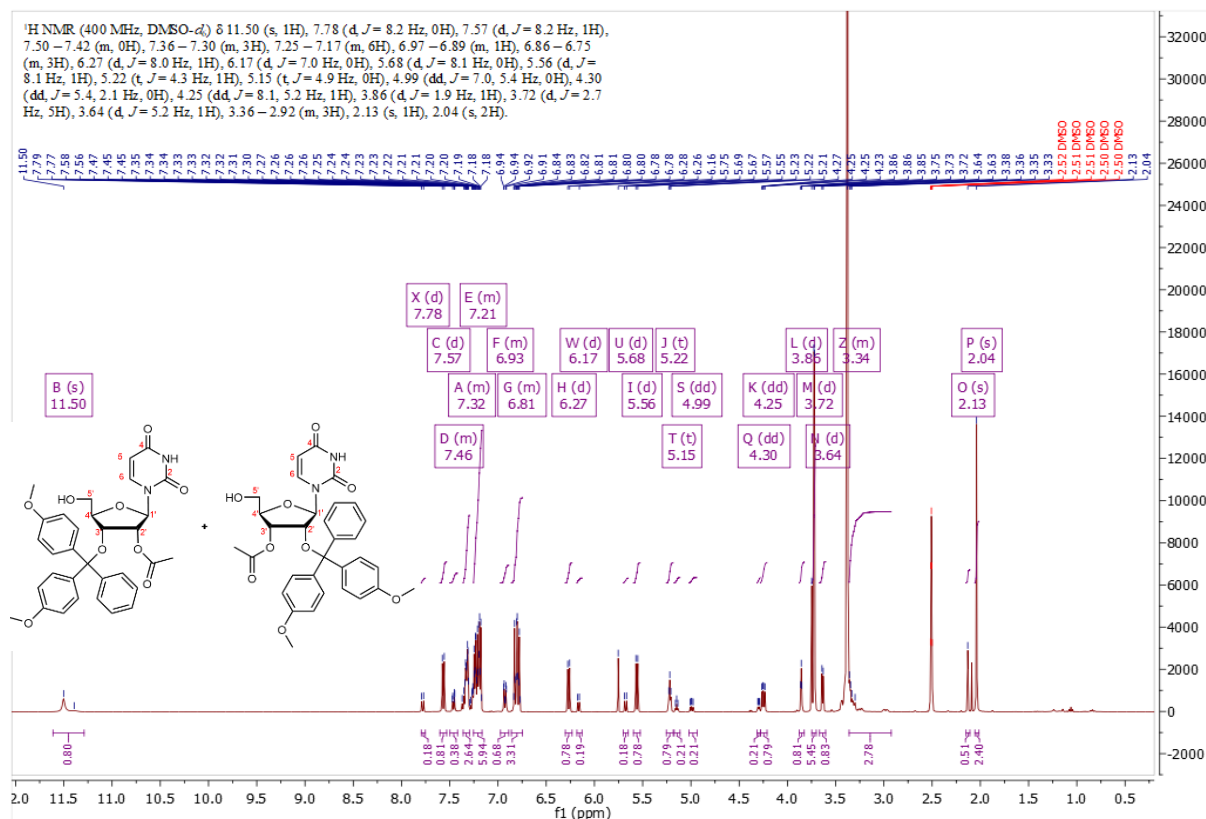


Figure S5.10 ¹H-NMR spectra of 2'-O-Acetyl-3'-O-(4,4'-dimethoxytrityl)uridine and 3'-O-Acetyl-2'-O-(4,4'-dimethoxytrityl) uridine (S5a+S5b)

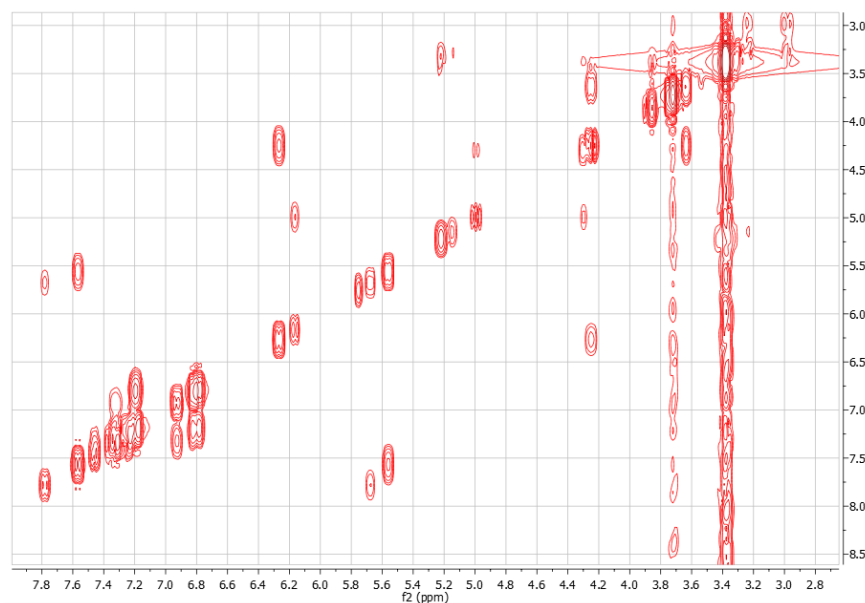


Figure S5.11 ¹H-¹H COSY NMR spectra of 2'-O-Acetyl-3'-O-(4,4'-dimethoxytrityl)uridine and 3'-O-Acetyl-2'-O-(4,4'-dimethoxytrityl) uridine (S5a+S5b)

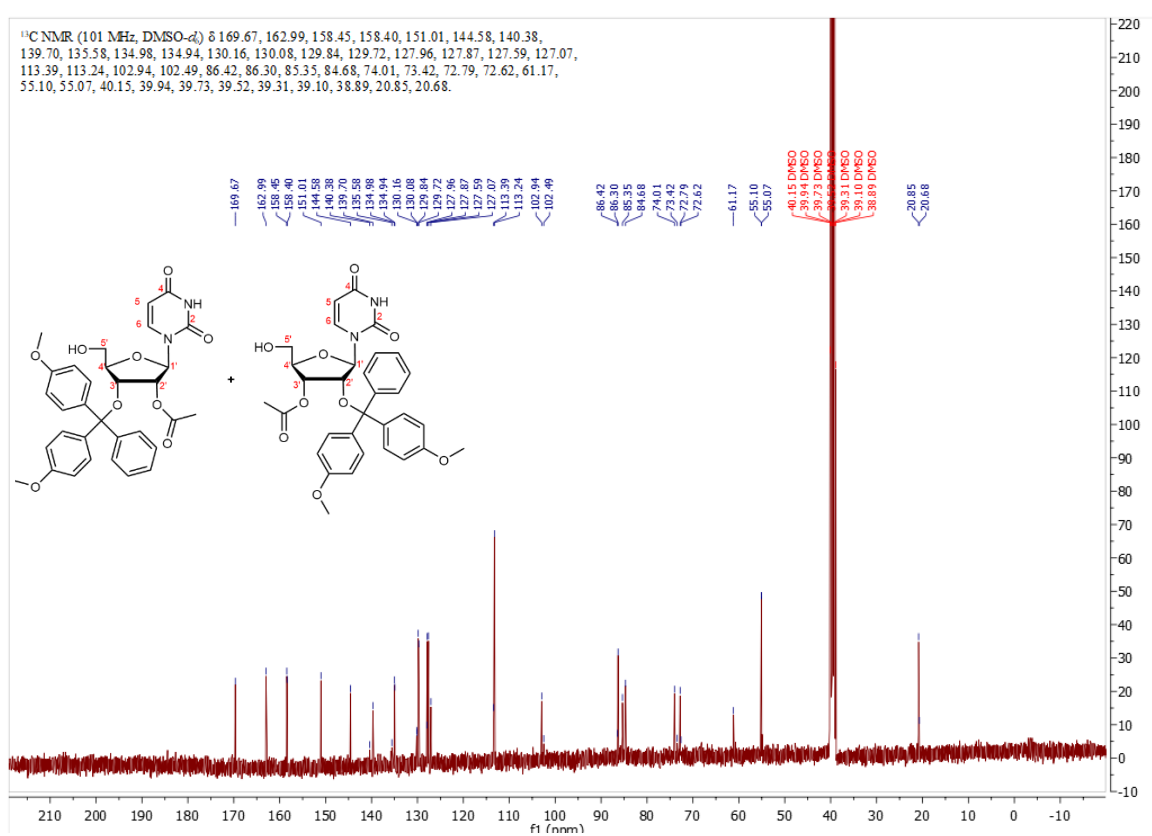


Figure S5.12 ¹³C-NMR spectra of 2'-O-Acetyl-3'-O-(4,4'-dimethoxytrityl)uridine and 3'-O-Acetyl-2'-O-(4,4'-dimethoxytrityl) uridine (S5a+S5b)

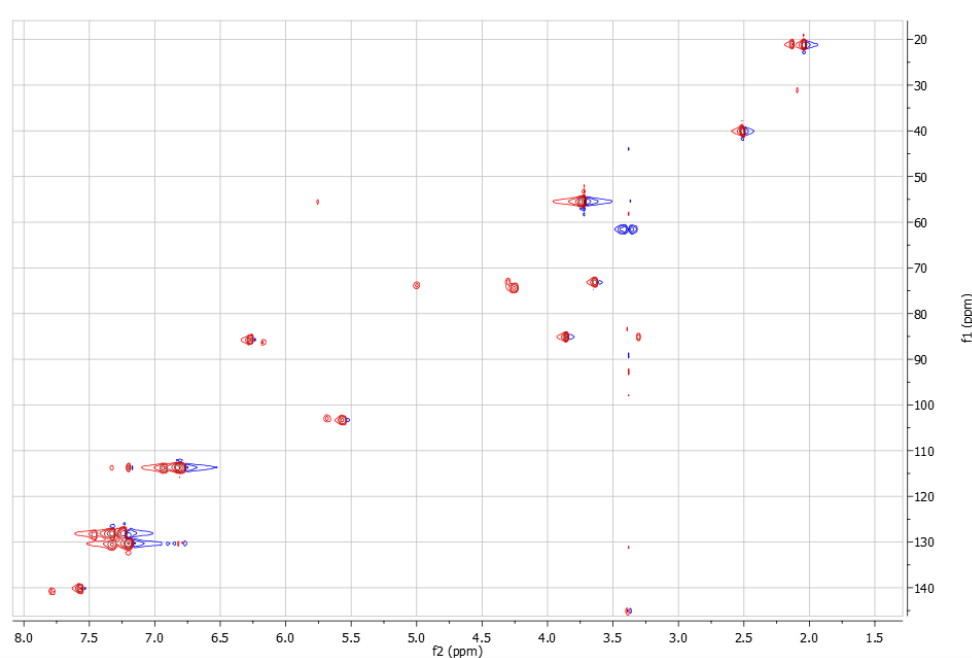


Figure S5.13 HSQC-NMR spectra of 2'-O-Acetyl-3'-O-(4,4'-dimethoxytrityl)uridine and 3'-O-Acetyl-2'-O-(4,4'-dimethoxytrityl) uridine (S5a+S5b)

5.5.1.6 5'-O-(N,N-Diisopropyl-2-cyanoethylphosphoramidite)-2'-O-acetyl-3'-O-(4,4'-dimethoxytrityl) uridine and 5'-O-(N,N-Diisopropyl-2-cyanoethylphosphoramidite)-3'-O-acetyl-2'-O-(4,4'-dimethoxytrityl) uridine (S6a+S6b)

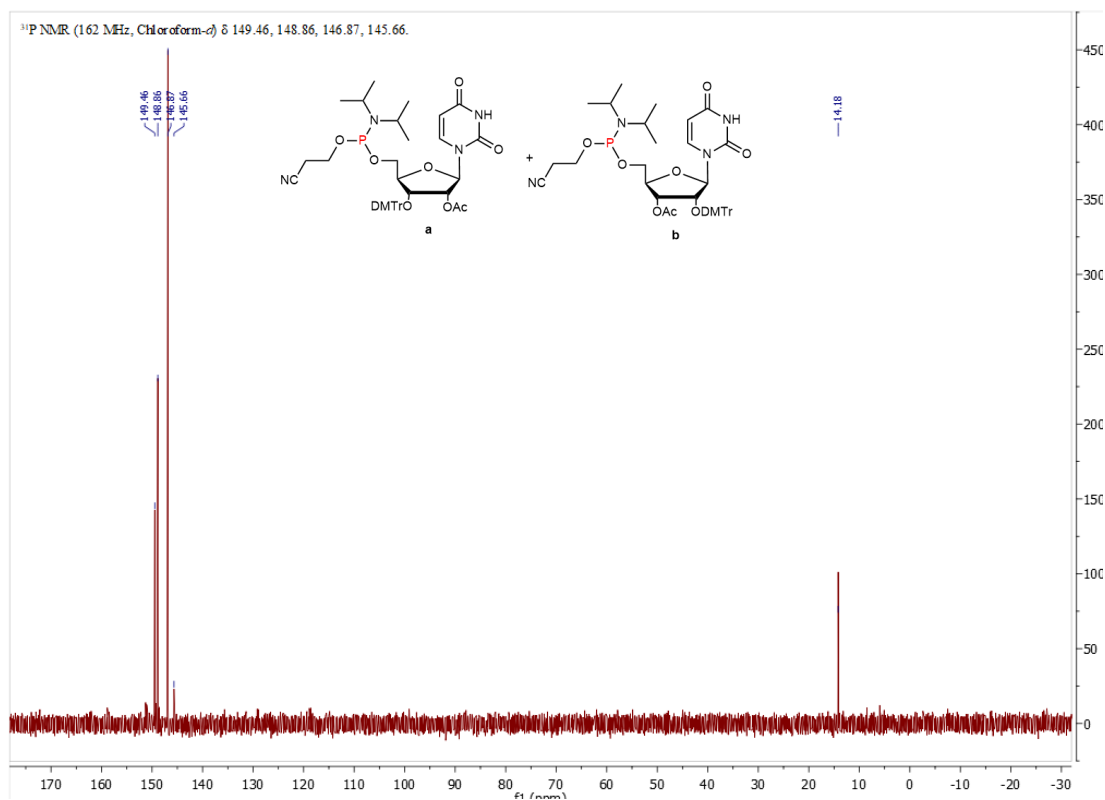
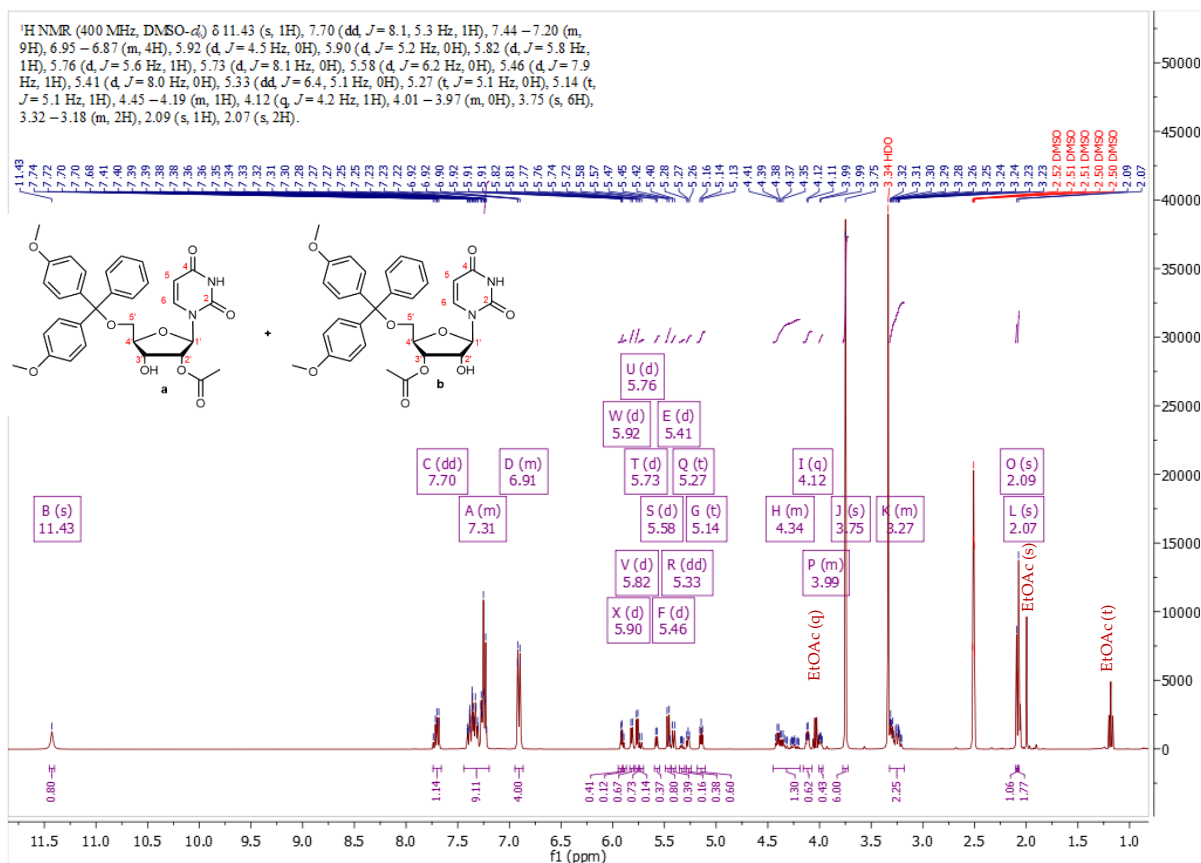
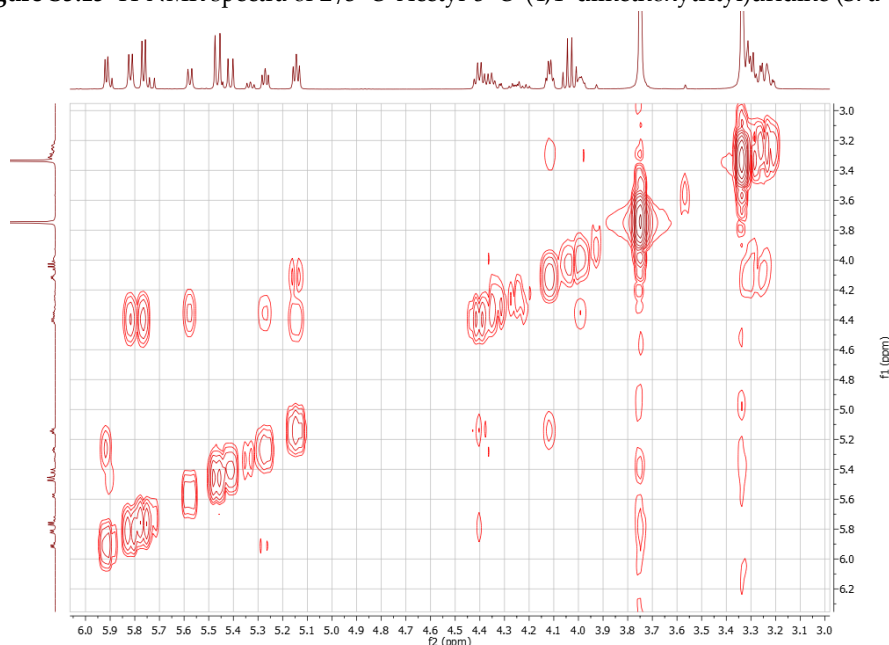


Figure S5.14 ^{31}P -NMR spectra of 5'-O-(N,N-Diisopropyl-2-cyanoethylphosphoramidite)-2'-O-acetyl-3'-O-(4,4'-dimethoxytrityl) uridine and 5'-O-(N,N-Diisopropyl-2-cyanoethylphosphoramidite)-3'-O-acetyl-2'-O-(4,4'-dimethoxytrityl) uridine (S6a+S6b)

5.5.1.7 2'/3'-O-acetyl-5'-O-(4,4'-dimethoxytrityl)uridine (S7a+S7b)

Figure S5.15 ¹H-NMR spectra of 2'/3'-O-Acetyl-5'-O-(4,4'-dimethoxytrityl)uridine (S7a+S7b)Figure S5.16 ¹H-¹H COSY NMR spectra of 2'/3'-O-Acetyl-5'-O-(4,4'-dimethoxytrityl)uridine (S7a+S7b)

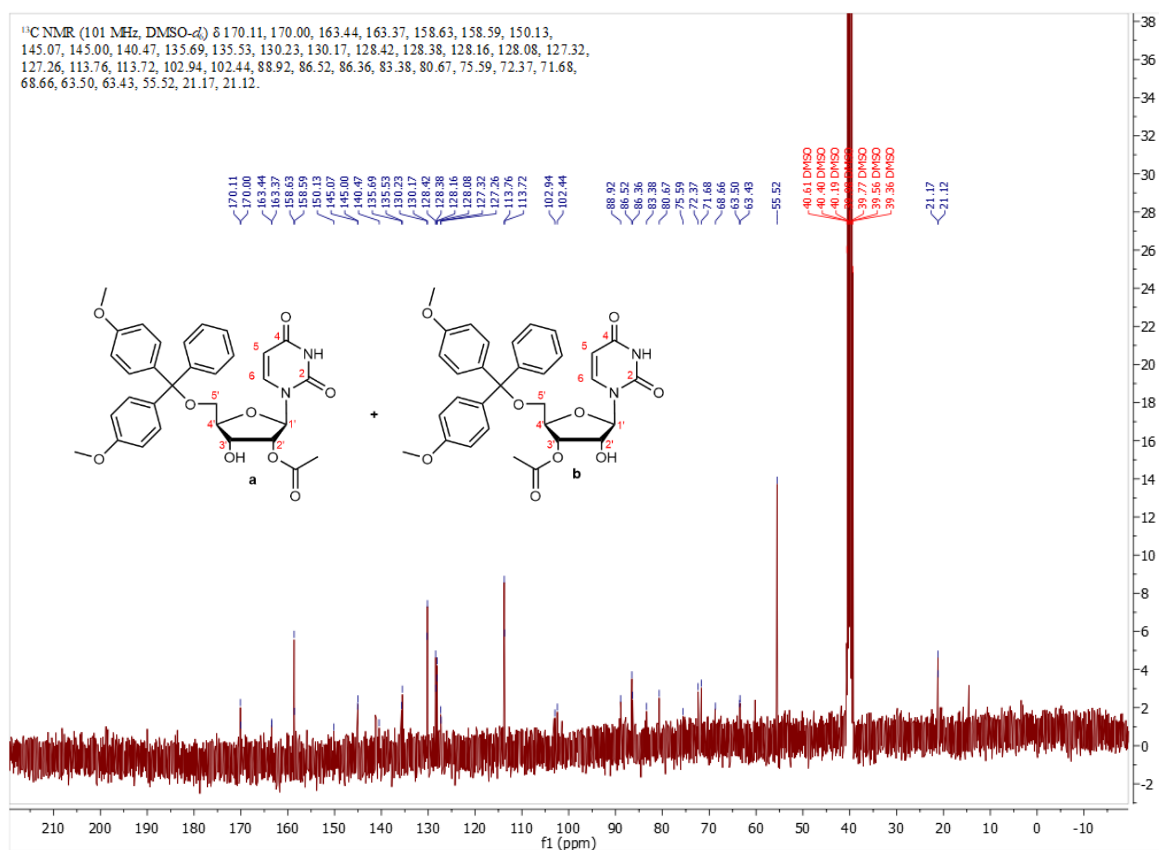


Figure S5.17 ^{13}C -NMR spectra of 2'/3'-O-Acetyl-5'-O-(4,4'-dimethoxytrityl)uridine (S7a+S7b)

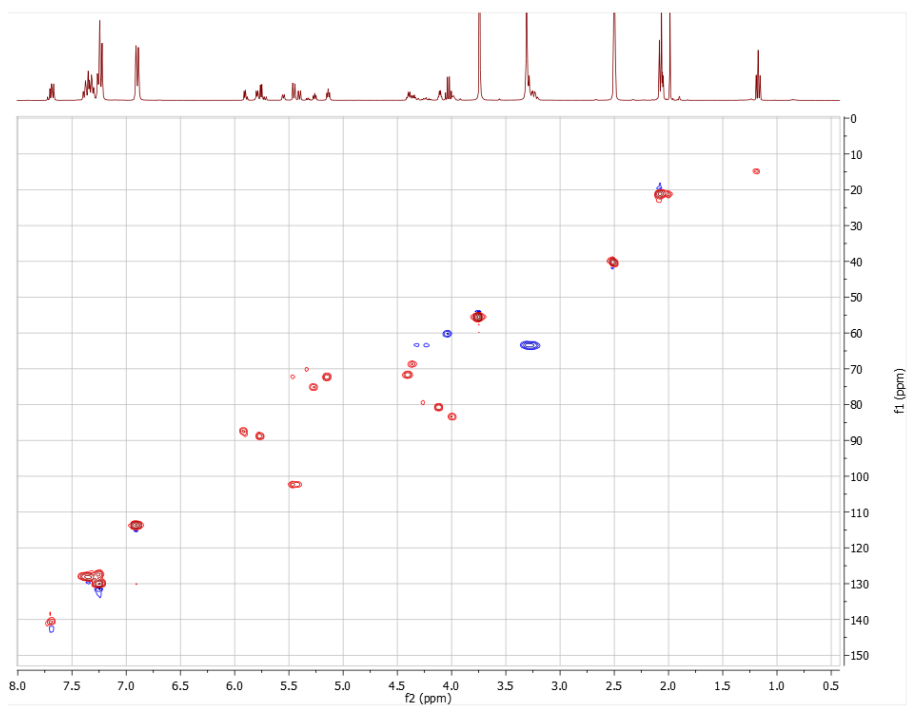


Figure S5.18 HSQC-NMR spectra of 2'/3'-O-Acetyl-5'-O-(4,4'-dimethoxytrityl)uridine (S7a+S7b)

5.5.1.8 2'/3'-O-Acetyl-5'-O-(4,4'-dimethoxytrityl)uridine-3'/2'-O-(2-cyanoethyl-N,N-diisopropyl)phosphoramidite (S8a+S8b)

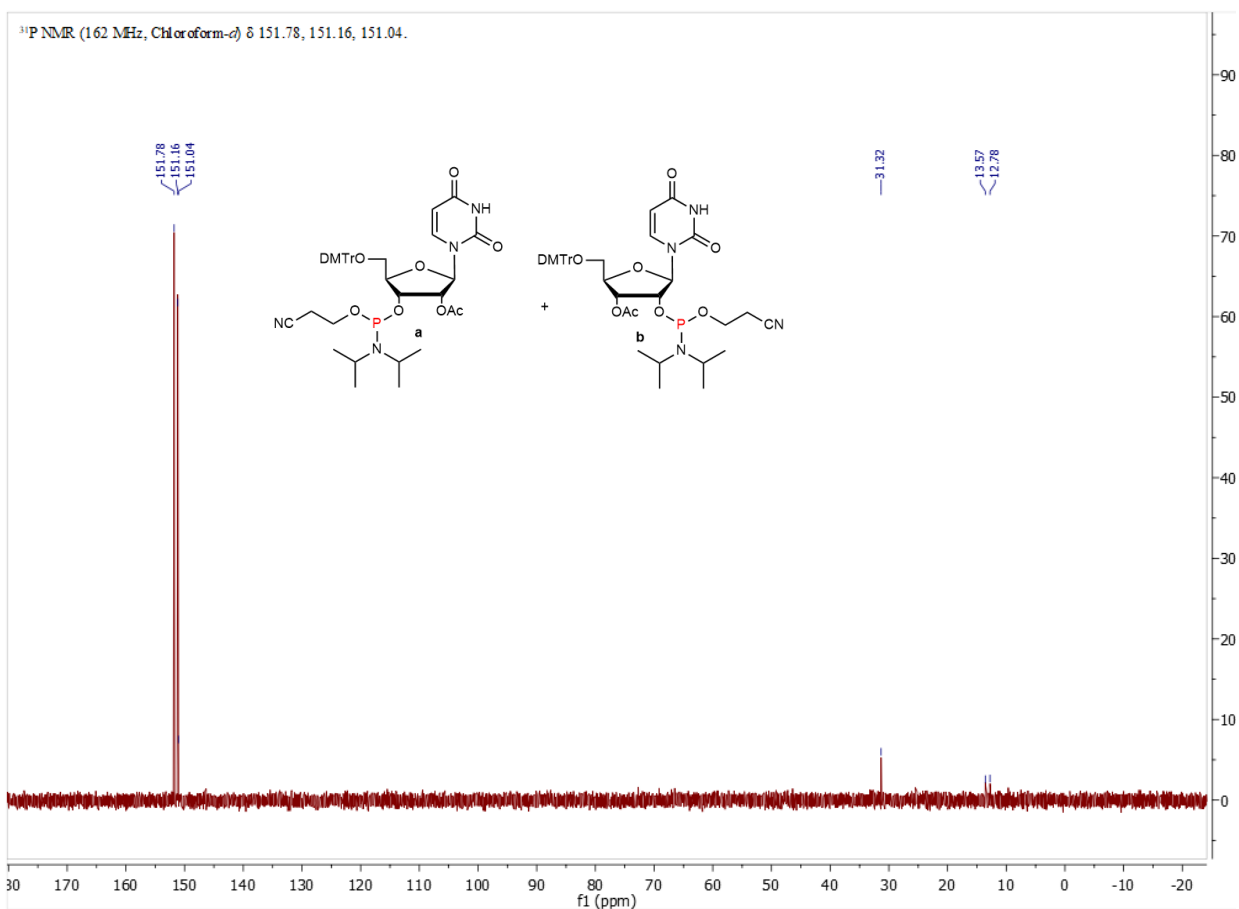
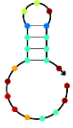
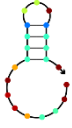



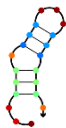
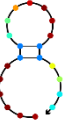
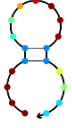


Figure S5.19 ³¹P-NMR spectra of 2'/3'-O-Acetyl-5'-O-(4,4'-dimethoxytrityl)uridine-3'/2'-O-(2-cyanoethyl-N,N-diisopropyl)phosphoramidite (S8a+S8b)

5.5.2 Oligonucleotide Design, Synthesis and Characterization

5.5.2.1 Design of microRNA

miRNAs were found in literature to have activity in breast cancer and then modified from sequences in miRBase (<https://www.mirbase.org/>) with 2'OMe in the 3'end, 5'end and middle to stabilize the sequence. Furthermore, miR27b-3p and miR129-1-5p were truncated with one and two nucleobases respectively. NUPACK (<https://nupack.org>) was applied to investigate theoretical hairpin structure at 37°C.

miR	Sequence 5'→3'	MW (g/mol)	Nt.	Regulation and Function in breast cancer cells	Hairpin (orig.)	Hairpin (modif.)
hsa-miR-31-5p	r(AGGCAAGAUGCUGGCAUAGCU)	6775.1	21	↓ Tumor suppressor: Involved in apoptosis, cell motility, cell invasion ¹⁷⁷		
Modified miR31	r(mAmGGCAAGmAmUGCUGGCAUAGmCmU)	6859.3	21			
hsa-miR-206	r(UGGAAUGUAA GGAAGUGUGUG G)	7202.4	22	↓ Tumor suppressor: Inhibit cell proliferation ¹⁷⁵		
Modified miR206	r(mUmGGAAUGmUmAAGGAAGUGUGUmGmG)	7286.5	22			
hsa-miR-27b-3p	r(UUCACAGUGGC UAAGUUCUGC)	6650	21	↑ Oncogenic: Promotes tumor progression and metastasis. Targets multiple tumor suppressors in breast cancer ¹⁸³		
Modified miR27-3p	r(mUmCACAGUG GmCmUAAGUUCUmGmC)	6428	20			
hsa-miR-129-1-5p	r(CUUUUUGCGG UCUGGGCUUGC)	6720.1	21	↓ Tumor suppressor: Inhibit proliferation and metastasis targeting CBX4 ¹⁷⁶		
Modified 129-1-5p	r(mUmUUUGCmGmGUCUGGGCUUmGmC)	6108.7	19			

Equilibrium probability




Table S5.1: microRNA table with original and modified sequences, biological function in breast cancer, MW, length (nt) and hairpin structure before and after modification.

5.5.2.2 Purchased control oligonucleotides

hsa-miR	MW (g/mol)	Sequences (5'→3')	nt
31 c	6859.3	mAmGrGrCrArArGmAmUrGrCrUrGrGrCrArUrArGmCmU	21
206 c	7286.5	mUmGrGrArArUrGmUmArArGrGrArArGrUrGrUrGrUmGmG	22
27b-3p c	6428.0	mUmCrArCrArGrUrGrGmCmUrArArGrUrUrCrUmGmC	20
129-1-5p c	6108.7	mUmUrUrUrGrCmGmGrUrCrUrGrGrGrCrUrUmGmC	19

Table S5.2: Purchased control miRNA from IDT. c = control

5.5.2.3 Trityl Values from Oligonucleotide Synthesis

a) P1 (KAS327)



b) P2 (KAS328)



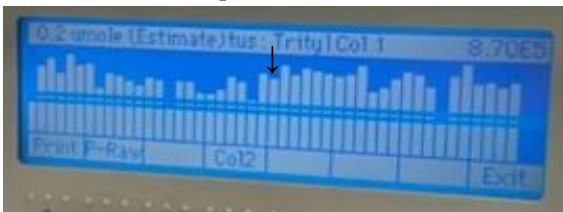
c) P3 (KAS329)



d) P4 (KAS330)



e) P5 (KAS331) part 1



f) P5 (KAS331) part 2



g) P6 (KAS332)

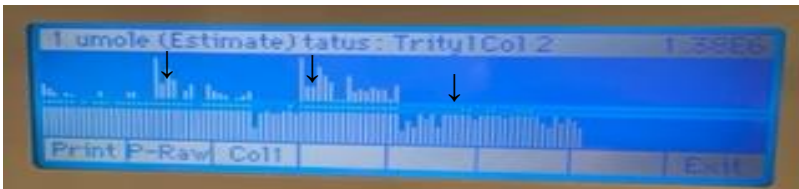


Figure S5.20. Trityl readout from oligonucleotide synthesis on Expedite Nucleic acid System. Note: Lines in the middle of the pictures are due to screen failure. Deprotection of branched linker S3 are shown as black arrows.

5.5.2.4 MALDI spectra of crude pools

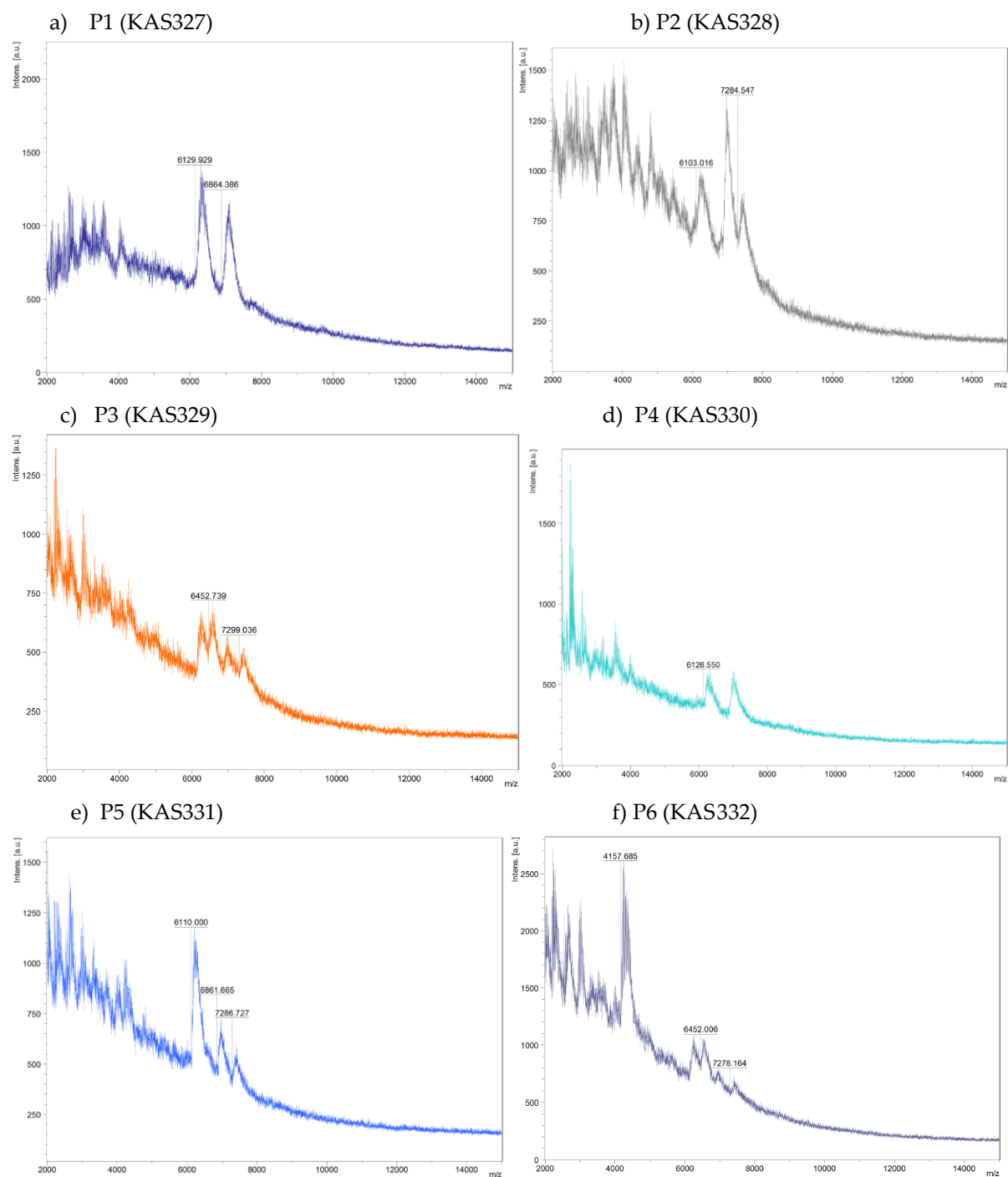


Figure S5.21. MALDI-TOF spectra of P1-P6 crude. Observed masses (m/z): (a) P1 (KAS327): 6129.93, 6864.39; (b) P2 (KAS328): 6103.02, 6850.84, 7284.55; (c) P3 (KAS329): 6109.14, 6453.91, 6862.93, 7296.97; (d) P4 (KAS330): 6126.55, 6864.44; (e) P5 (KAS331): 6110.00, 6861.67, 7286.73; (f) P6(KAS332): 4157.69 (failure), 6122.13, 6452.01, 6840.42, 7278.16.

5.5.2.5 Standard curves of miRNA controls

Dilution row was made from having 60µl of stock concentration (800ng/µl) in vial 1, and transferred 40µl to vial 2 containing 20µl and mixed 10 times, before transferring the next 40µl to vial 3 and so on. This resulted in a dilution factor of 1.5 with 20µl in each vial for HPLC. 10µl was injected and run at 75°C, 5-15% over 30min. Values from HPLC was written into lab notes and standard curves constructed.

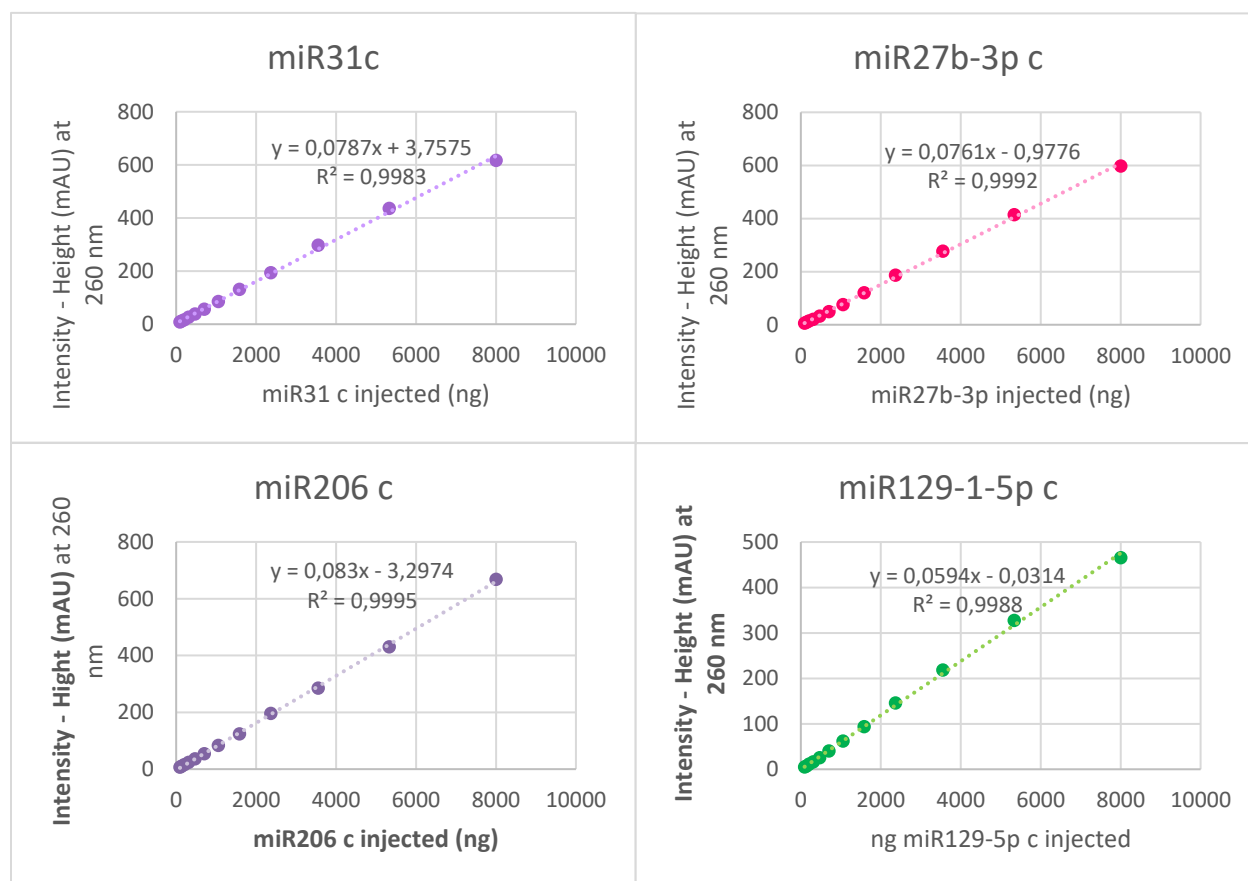


Figure S5.22: Standard curves for control miRNA: miR31, miR206, miR27b-3p, miR129-1-5p.

5.5.2.6 Yield estimation of miRNA in pools

Yields and amount of each miRNA in each pool was calculated from standard curve equations shown above along with purity of each HPLC peak. miR31 and miR206 have the same retention time, so these were calculated as a mix or one peak in P2, P3, P5, P6. Pools were made on 200nmol scale and was after purification dissolved in 650µl RNase free water. 10µl of each pool were injected to HPLC and UV readout (mAU) of each peak was evaluated relative to standard curves.

Pool	miRNA	Eq.	mAU (hight)	Purity (%)	ng injected	%Volume injected	ng/ul	%Total volume (650 ul)	ng (total)	nmol (total)	Estimated Yield%
P1	129-5p	1	138.72	27.6	2323.76	2.76	841.94	179.4	151044.5	25	24.7 %
	31	1	205.1	38.3	2558.36	3.83	667.98	248.95	166293	24	24.2 %
P2	129-5p	1	67.04	19.95	1129.15	1.99	565.99	129.68	73394.63	12	18.0 %
	31	1	127.75	44.67	1575.51	4.47	352.70	290.36	102408	15	11.2 %
	206	1									
P3	129-5p	1	75.04	16.13	1263.82	1.61	783.53	104.85	82148.84	13	26.9 %
	31	1	145.59	37.5	1802.19	3.75	480.58	243.75	117142.5	17	17.1 %
	206	1									
	27b-3p	1	75.04	16.13	1263.83	2.07	783.53	104.85	82148.84	13	26.9 %
P4	129-5p	2	68.05	31.95	1146.15	3.20	358.73	207.68	74499.85	12	6.1 %
	31	1	187.4	63.38	2333.45	6.34	368.17	411.97	151674.2	22	22.1 %
P5	129-5p	4	178.8	28.49	3010.63	2.85	1056.73	185.19	195690.9	32	8.0 %
	31	2	234.02	48.91	2925.83	4.89	598.21	317.92	190178.7	28	9.2 %
	206	1									
P6	129-5p	8	85.58	19.8	1441.27	1.98	727.92	128.7	93682.51	15	1.9 %
	31	4	110.67	28.15	1358.48	2.82	482.59	182.98	88301.3	13	2.1 %
	206	2									
	27b-3p	1	131.15	25.01	1736.24	2.50	694.22	162.57	112855.4	18	17.6 %

Table S5.3: Estimated amount and yield of each miRNA in each pool based on calculations from standard curves

6 PCR-Free Targeted Single Primer Capture for Ultra-Deep Sequencing

This chapter is about the experiments conducted at Stanford University from October 2021 to March 2022 as a part of my external stay. It concerns the optimization of the Oligonucleotide-Selective Sequencing method on APEX beads (OS-Seq-APEX) using branched oligonucleotide as primer spacing for improving capture performance. The experiments were carried out in the Ji Research Group, Division of Oncology at Stanford University with supervision from GiWon Shin and Professor Hanlee P. Ji.

6.1 Oligonucleotide-Selective Sequencing (OS-Seq)

Targeted sequencing has shown to be an important tool in cancer diagnostics^{28,114}. Target methods reduce the cost and time for Next-Generation Sequencing, by capturing genomic regions of interest from a DNA sample before sequencing. Several targeted sequencing methods have been developed^{115,184–186}. Oligonucleotide-selective sequencing (OS-Seq) is a targeted sequencing technique developed in the Ji Research Lab that involves capturing and sequencing genomic target DNA on a sequencers solid-phase, such as the Illumina flow cell^{187,188}. This method was developed to overcome limitations faced using other targeted sequencing techniques at the time^{115,187,188}. However, due to less freedom modifying existing Illumina platforms, the OS-Seq was moved from flow cell to agarose beads, and currently needs to be optimized. The first generation on flow cell will be referred to as *OS-SeqV1*, second generation of agarose beads *OS-SeqV2* and optimizations of this *OS-SeqV2.1*.

6.1.1 Principals of OS-Seq

The OS-Seq method can be divided into three overall steps (**Figure 6.1**): 1) Flow cell modification step, where flow cells are modified with primer-probes containing target-probe (purple) and sequencing primers (green). 2) Library extension step, where desired targets are captured by the immobilized primer-probes. 3) Lastly, captured library are sequenced using standard sequencing protocol to generate *Read 1* and *Read 2* holding the genomic information of the targeted capture sequences.^{187,188}

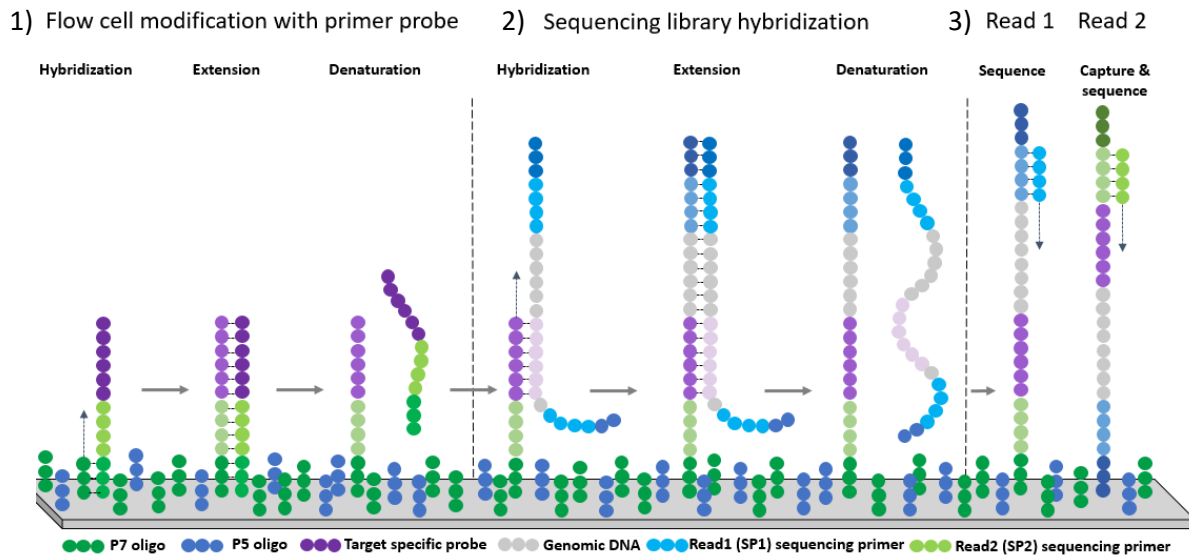


Figure 6.1: First generation of oligonucleotide selective sequencing (OS-Seq)^{187,188}. 1) Flow cell modification with primer-probe; 2) Library hybridization and extension; 3) Sequencing¹⁸⁸.

In the first generation of OS-SeqV1, the existing P7 primer on Illumina flow cell is 1) modified to become a target-specific primer probe. This is done, by hybridizing with its complement primer probe. Primer probe contain target specific sequence (purple) and sequencing primers SP2 (lime green) and P7 (green) in the design (**Figure 6.1**). By extending the P7 primer with polymerase followed by denaturation, a set of randomly placed target-specific primer probes are generated on the flow cell surface. Next, 2) the extended primer probe capture target sequence of prepared DNA library and function as a primer for polymerase extension. The library is prepared with adaptors that enable binding with primers on flow cell and multiplexing. After hybridization, extension and denaturation, the 3) extended library is bridge amplified to make clusters of library molecules on the flow cell ready sequencing (Read 1 and Read 2). The captured molecules are bridge amplified and sequenced using standard sequencing protocols¹⁸⁷.

The first generation OS-Seq was applied to resequence exons of either 10 or 344 cancer genes from human DNA samples. This showed selectivity >87% of captured sequences originated from the intended target regions¹⁸⁷. Sensitivity >95% of single nucleotide variants called with OS-Seq agreed with whole-genome sequencings (WGS) for the same individuals¹⁸⁷. Later on, the OS-Seq method was further improved by Hopmans et al. where all steps, except for library preparation, were automatized to eliminate complicated sample preparations steps¹⁸⁸.

6.1.2 Second Generation of OS-Seq on Agarose Beads

Recently, Illumina made their sequencing platforms more user-friendly, limiting the ability and degree of freedom to modify the flow cell. Therefore, in the second generation (OS-SeqV2), the chemistry was moved from flow cells to APEX (attachment-based primer extension) magnetic agarose beads functionalized with tetrazine (Tz). APEX beads has previously been used to attach genomic library allowing for iterative molecular analysis and storage¹⁸⁹. The workflow in OS-SeqV2 is illustrated in **Figure 6.2A-E**, and the steps described in the section below.

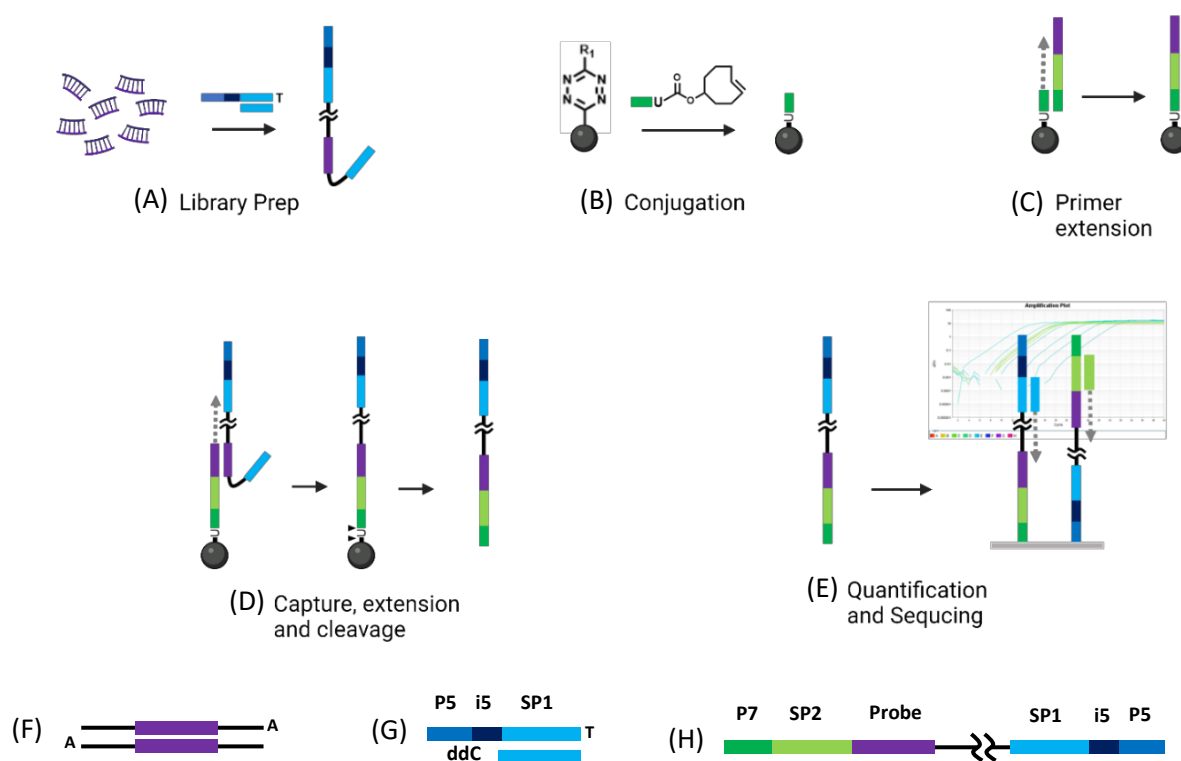


Figure 6.2: Workflow in OSSeqV2 method: (A) Library prep of genomic DNA. (B) Conjugation of primer probe (P7) to APEX beads. (C) Primer extension of P7 with target probe pool. (D) Library capture, extension and cleavage of final extended library. (E) Quantification, amplification and sequencing of library. (F) Fragmented dsDNA with A-tails; (G) Applied adapters; (H) Cleaved library from beads ready for sequencing.

Library preparation. Genomic DNA (NA12822, Coriell Institute) were first fragmented to size of ~500nt in average using fragmentation enzymes, and then treated with End-Repair and A-tailing enzymes followed by ligase so the fragmented DNA could be ligated to adaptors (**Figure 6.2A+F**). The adaptors contained P5-primer, 5index (unique barcode on 8nt) and SP1-primer all needed for sequencing and identification (**Figure 6.2H** and **Table S6.2**). The library is prepared for a Single-Indexing Paired-End Flow Cell Sequencing.

Primer conjugation to APEX beads. By functionalizing the P7 probe with *trans*-cyclooctene (TCO) the P7 primer can be conjugated to magnetic beads through an inverse electron demand Diels–Alder cycloaddition (iEDDA) (**Figure 6.3**)^{190,191}. This reaction occurs between an electron-rich dienophile (alkene in TCO) and an electron-poor diene in Tz to make a [4+2] cycloaddition, followed by a retro [4+2] cycloaddition to liberate N₂. This reaction will generate a 4,5-dihydropyridazine species which either isomerize to 1,4-dihydro-isomers or oxidize to form pyridazine product¹⁹⁰. iEDDA click reactions has emerged into biology, imaging and therapeutic applications due to its fast reaction kinetics, excellent orthogonality and biocompatibility allowing *in vivo* reactions¹⁹¹.

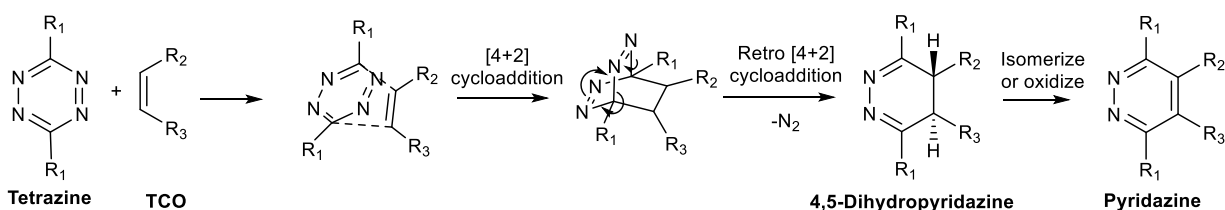


Figure 6.3: Mechanism of iEDDA: Schematic representation of reaction between tetrazine (Tz) and *trans*-cyclooctene (TCO)^{190,191}.

Primer Extension. When conjugated to beads, the P7 primer was extended with different target probes. This is a pool of different oligonucleotides on 101nt in size, containing complement P7 (24nt, green) and complement SP2 (37nt, lime green) and targeting sequences (40nt, purple), all needed for capturing and sequencing (**Figure 6.2C**). In this project, we have used OS42 consisting of 6205 different oligonucleotides targeting 84 different oncogenes relevant for colon cancer¹⁹². When the oligonucleotide template has hybridized to conjugated primer, polymerase (Q5U) is added to extend the primer and the OS42 is denatured so only the extend primer remains on the beads.

Library Capture, Extension and Cleavage. The prepared library is added to the beads containing extended primer probe ensuring the target probe can catch targeted region in genomic library. The library is extended on the beads using *Bst* DNA polymerase to obtain a full library on beads with different genomic target sequences of interests. After library capture, the library is cleaved from the beads with USER enzyme. The USER enzyme is an uracil glycolase enzyme that recognizes dU in the sequence by removing the base. The resulting structure is not stable so it will readily undergo first β -elimination followed by δ -elimination to release the two phosphate-functionalized oligonucleotides and α,β -unsaturated aldehyde as byproduct (**Figure 6.4**)¹⁹³.

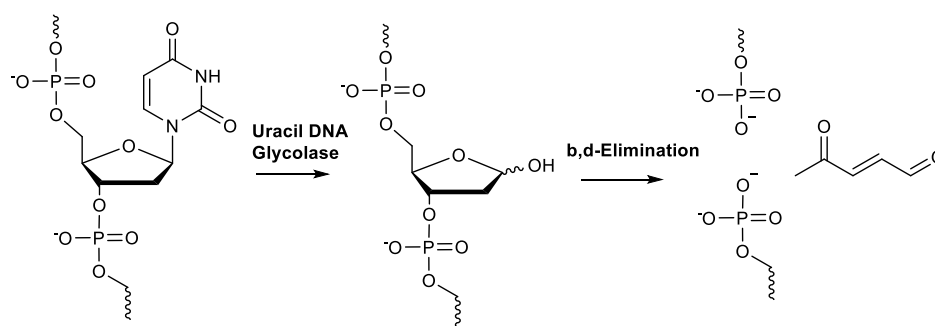


Figure 6.4: Mechanism for USER enzyme (Uracil DNA Glycolase) removing uracil, followed by spontaneous β,δ – elimination to liberate the two oligonucleotides and a α,β -unsaturated aldehyde¹⁹³.

Quantification and Sequencing. The cleaved library is quantified using qPCR followed by amplification with PCR. Ideally, the method should be PCR-free, however in order to have enough material and ensure high sensitivity and selectivity, PCR is applied. After PCR amplification, the library is quantified again with qPCR and sequenced using iSeq 100. Reference genome (PhiX) is added to align the reads correctly. Sequencing is made on a Paired-End Flow cell from Illumina (**Figure 1.5B**) with 300 cycles meaning, SP1 primer is added, and 150 bases are read in the forward direction (genomic part of library). Next, the adapter in library bind to P5 primer on the flow cell initiating i5 reading (8nt barcode). The library is bridge amplified, and SP2 primers are added, such 150 bases are read from the other side of the library but in forward direction (targeted + genomic part of library).

6.1.3 Sequencing parameters

Post sequencing, the data was evaluated and analyzed using *in-house* bioinformatics pipeline created in the Ji Group specific for the OS-Seq method^{187,188}. The pipeline is only applied as a tool and not discussed in the thesis. To compare the performances of the first and second generation of OS-Seq we looked at three main parameters: *Total number of reads*, *%On-target reads*, and *%OK reads*. Total number of reads is important in terms of sensitivity. The reads are designed to cover specific target regions in the genome. Therefore the percentage of aligned reads covering the target area (%On-target) is also important factor and relates to selectivity. Lastly, %OK reads, is the number of reads (*read 1* and *read 2*) that lays within the target region and do not overlap. **Table 6.1** shows the performance of first generation (V1) and second generation (V2) of OS-Seq with current conditions optimized by GiWon Shin (GS). It is seen that V2 does not yet perform in the same way as V1 in terms of number of reads on target (82% vs. 20%) and OK reads (43% vs. 9.5%). Therefore, the current platform on APEX beads needs to be optimized in order to be suitable for targeted sequencing in diagnostic cancer samples.

OS-Seq Parameters	V1 (Run633, L006942) ¹⁸⁸	V2 (SP1 blocker, 1/4 beads)*
%On-target	82%	20%
On-target read pair	8,325,559	110,218
Ok reads	7,146,544	20,978
%Ok	42.92%	9.52%

Table 6.1: Current performance of first (V1) versus second (V2) generation of OS-Seq.

*Numbers of V2 obtained by GiWon Shin and reprinted with permission (unpublished).

6.2 Project outline: Optimization of OS-SeqV2

To optimize the current Oligonucleotide-Selective Sequencing method (OS-SeqV2) we hypothesize the amount and quality of reads can be increased by improving capture performance. Capture performance is defined, by how well the extended primer probe bind end elongate the prepared genomic library. There are different approaches to optimize these conditions. Looking at the primer-library intermediate before extension (**Figure 6.5**), one could imagine several things interfering and affecting the performance of the elongation. The SP1 primer in the 5'-end, originating from the library preparation, could potentially self-hybridize to complementary part in the 3'-end thereby inhibiting polymerase activity. Likewise, unspecific binding of SP2-primer or non-extended P7-primer could also result in shorter reads or reads missing important parts. To avoid incorrect binding and interference, different blockers (SP1-blocker, SP2-blocker and P7-blocker) could be added prior to elongation. Another target for optimization could be varying the space between the P7-primer on the beads to improve conditions for the enzyme to function.

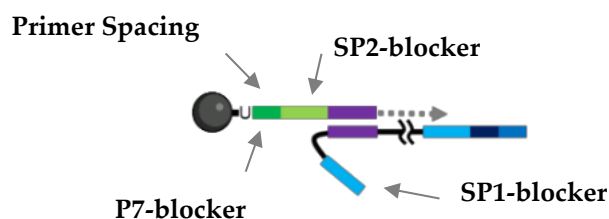
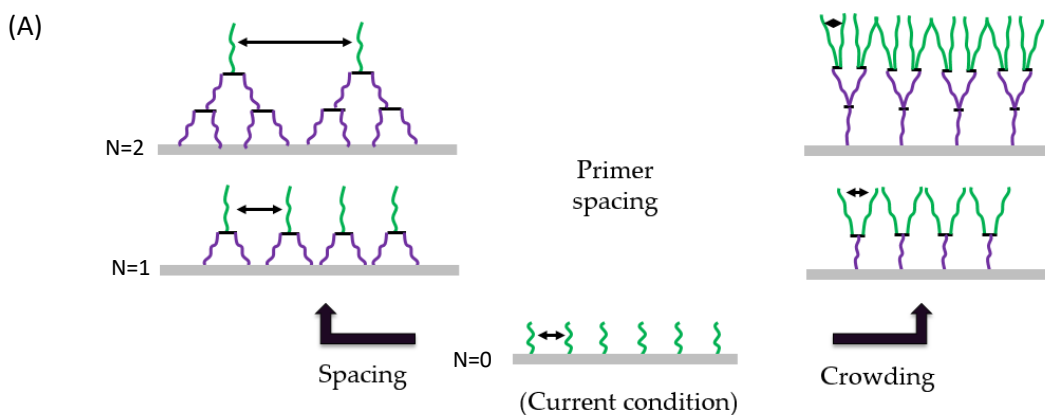


Figure 6.5: Potential optimization strategies for improving capture performance.

The overall goal of this project was to improve existing OS-SeqV2 platform by optimizing capture performance, and ideally make it PCR-free. Therefore, we designed six different branched oligonucleotides with different spacing (**Figure 6.6A**). By increasing or decreasing the distance between P7 primers on the APEX beads using branched oligonucleotides instead of conventional linear primers, we could potentially improve capture performance. Branched oligonucleotides have previously been applied for signal amplification in branched assays for use in single cell sequencing¹⁹⁴. Furthermore, cone-shaped dendrimers have been used for spacing in microarrays to improve yields of detected DNA and improved discrimination^{117,118}. Therefore, we decided to make branched oligonucleotides for spacing. The effect of different primer lengths and blockers were also investigated. We looked at SP1 blocker, P7 blocker, SP2 blocker and extended P7-primer (P7.2) containing a phosphorothioate (PS) bond in the 3'end as potential factors to improve capture performance and library yield.

6.2.1 Design of branched oligonucleotides

We designed six branched oligonucleotide that could generate more and less space between the primers on the APEX-beads. *Spacing* branched oligonucleotides (Sb) for increased space and *crowding* branched oligonucleotides (Cb) for decreased space (**Figure 6.6A**). It was assumed that Sb-primers would improve capture performance, but to have the opposite effect to compare with, Cb structures were also designed and synthesized.



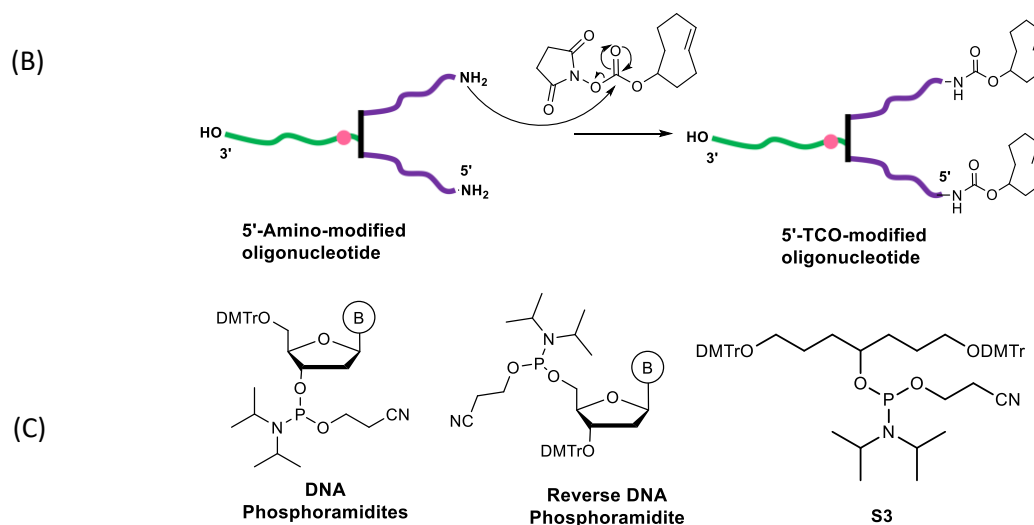


Figure 6.6: Principals of primer spacing using branched oligonucleotides. (A) Sb-primers (increased space) and Cb-primers (decreased space); (B) NHS-coupling of amino-modified oligonucleotides with TCO-NHS-ester to obtain TCO-primers for OS-SeqV2; (C) Types of phosphoramidites applied in synthesis of Cb and Sb oligonucleotides.

The Sb structure was designed so the functional primer sequence (P7, green in **Figure 6.6A**) would point outwards with 3'-OH group, enabling enzymatic reactions and having the 5'-end anchoring the APEX beads. Sb-structures were synthesized from the 3'-end starting with the P7 sequence with incorporation of a deoxy-uridine base (dU) enabling cleavage from beads, and a T₉-spacer. From the spacer, the branching linker phosphoramidite (S3) was added, from where two secondary sequences (RS1) were made in two branches, each with 5'-amino-modifiers in the end. After synthesis, the amino-modified ends were conjugated with TCO-NHS-ester enabling conjugation to the Tz-beads (**Figure 6.6B**).

Likewise, the Cb structures were “flipped” in structure compared to Sb-structures. Consequently, the RS1 would anchor the APEX beads, and the 5'-ends would have P7 primer branches crowding the space. However, because enzymes only work from 3'-end starting with a hydroxyl or phosphate group, we built the P7 primer branches from dU in reverse direction with reverse phosphoramidites (**Figure 6.6C**). The 3'-amino-modified RS1 end was conjugated with TCO-NHS ester to enable Tz-bead conjugation.

We also wanted to investigate numbers of branches ($N = 0, 1, 2 \dots n$) in spacing and crowding (Sb1 and S2b1, Cb1 and C2b1), the length of primer sequence (P7.1 and P7.2, Sb2.1 and Sb2.2) and the length of secondary probe (Sb2.1 and Sb2.2). In total, we made six different branched TCO modified oligonucleotides: Four spacing and two crowding branched oligonucleotides for use in the OSSeqV2. The TCO-P7.1 and TCO-P7.2 primers were applied as control (**Figure 6.7**).

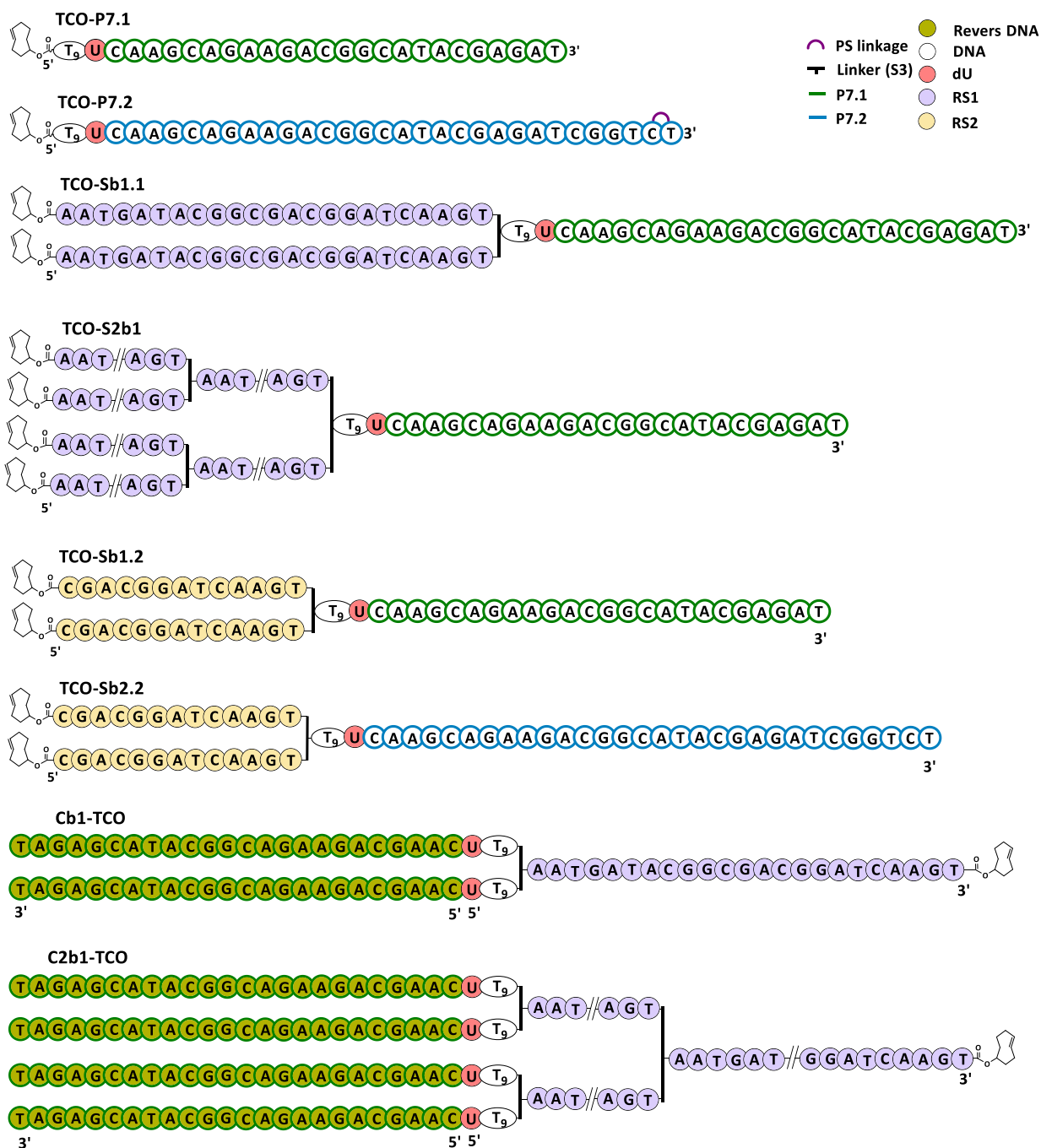


Figure 6.7: Chemical structures of TCO-Primers tested in OS-SeqV2 method. Branched oligonucleotides designed and synthesized for optimization study of OS-SeqV2: TCO-P7.1; TCO-P7.2, TCO-Sb1, TCO-S2b1, TCO-Sb1.2, TCO-Sb2.2, Cb1-TCO and C2b1-TCO. Sb: Spacing; Cb: Crowding; S2b1/C2b1: N=2. TCO-Primers with P7.1 sequence are green, P7.2 sequence are blue; RS1: random sequence 1 (violet); RS2: random sequence 2 (yellow); DNA (white) and revers phosphoramidites (green); PS linkage (purple); dU (pink).

6.3 Results and Discussion

6.3.1 Synthesis of Branched Oligonucleotides

Branched oligonucleotides were synthesized using solid-phase synthesis with DNA phosphoramidites. Branched linker **S3** was synthesized and applied to obtain branched structures (**Figure 6.6C**). Adrian Hernández Bustos helped with synthesis of the final structures. We tried to obtain MS spectra using MALDI-TOF but without success as the branched oligonucleotides did not fly. Consequently, the oligonucleotides were not tried purified with HPLC. Instead, the oligonucleotides were tried characterized with Novex™ denaturing polyacrylamide TBE-Urea gels (1xTBE, 7M Urea, 10%) (**Figure 6.8A**).

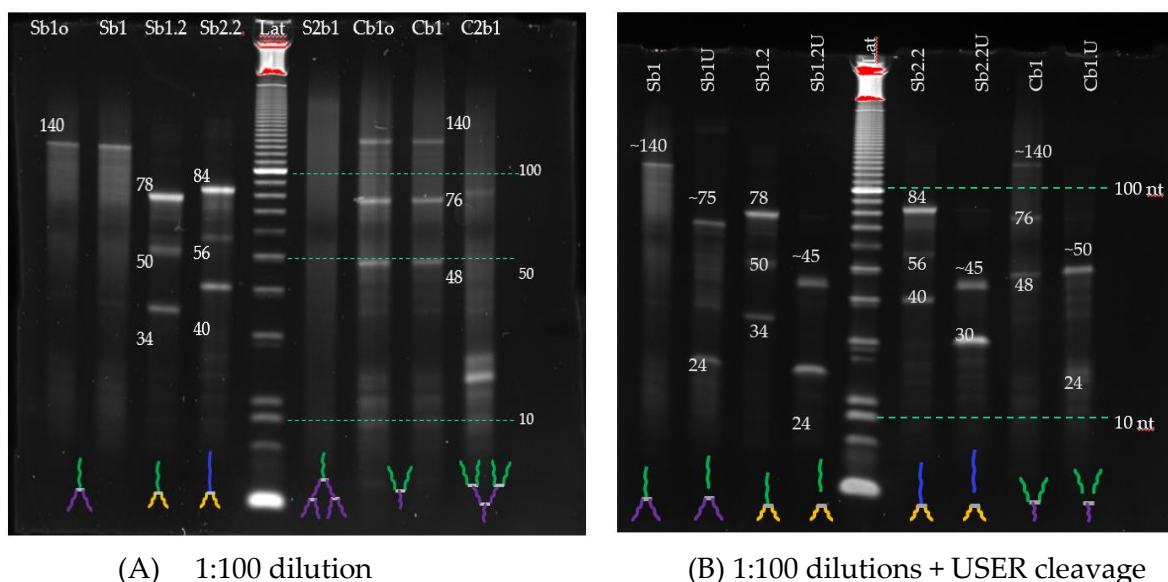


Figure 6.8: PAGE of TCO-branched primers (Sb1, Sb1.2, Sb2.2, S2b1, Cb1 and C2b1) (left), cleaved with USER enzyme (Sb1, Sb1.2, Sb2.2, Cb1) and cleaved Sb1U, Sb1.2U, Sb2.2U, Cb1U. 1:100 dilutions in 1xTBE, 7M Urea, 10% PAGE.

It was expected, the branched structure might act differently than linear sequences in the gel, and have different running patterns. From **Figure 6.8A** it is seen that Sb1 (in lane 1 and 2) gave smears with one band around 140nt which is much above the expected linear length of 59nt (80nt in total). Furthermore, it was observed that Cb1, Sb1.2 and Sb2.2 all gave a three-band pattern with the strongest signal on top (much higher than expected) and weaker bands below, with similar difference in between. S2b1 and C2b1 gave only smears probably because they were more branched than Sb1 and Cb1.

Because of the unexpected pattern in the gel, we could not confirm the size of the oligonucleotides. However, we could confirm they were not running as linear structures, and therefore could be branched. We decided to treat a small portion of Sb1, Sb1.2, Sb2.2 and Cb1

with USER to cleave each oligo into two fragments (**Figure 6.8B**). Sb1 would give two fragments: a linear at 24nt and branched at >57nt that might lay higher due to the structure. Similar for Sb1.2: 24nt + >37nt; Sb2.2: 30nt + >37nt; Cb1: 24nt + >42nt. Using this approach we could confirm the size of the structures (**Table 6.2**). After cleavage with USER two bands appear for each oligo, all showing the right linear primer sequence length at 24nt or 30nt, with a secondary fragment laying higher than calculated due to branched structure. This confirmed the size of Sb1, Sb1.2, Sb2.2, Cb1. Testing of S2b1 and C2b1 will be done in near future. We decided to continue with TCO-NHS coupling and use the oligonucleotides in assay.

Name	Sequence 5'→3'	Length [nt] (total)	MW [g/mol]	Length obs. [nt]	Length obs.+ USER [nt]
P7.1	/5TCO-PEGN/TTTTTTTTTUC AAGCAGAAGACGGCATACGAGAT	34	11035.5	-	-
P7.2	/5TCO-PEGN/TTTTTTTTTUC AAGCAGAAGACGGCATACGAGATCGGTC*T	40	12896.7	-	-
Sb1	5'-[TCO]-Y-AATGATACGGCGACGGATCAAGT-X- TTTTTTTTTUC AAGCAGAAGACGGCATACGAGAT-3'	59(80)	25285.6	140	24+ ~75
S2b1	5'-[TCO]-Y-AATGATACGGCGACGGATCAAGT-X- AATGATACGGCGACGGATCAAGT-X- TTTTTTTTTUC AAGCAGAAGACGGCATACGAGAT-3'	83(172)	54584.9	smear	-
Sb1.2	5'-[TCO]-Y-CGACGGATCAAGT-X- TTTTTTTTTUC AAGCAGAAGACGGCATACGAGAT-3'	49 (63)	19009.4	34,50, 78	24+ ~45
Sb2.2	5'-[TCO]-Y-CGACGGATCAAGT-X- TTTTTTTTTUC AAGCAGAAGACGGCATACGAGATCGGTCT-3'	55 (69)	20854.6	40,56, 84	30+ ~45
Cb1	3'-tagagcatagcagaagaacgaac-5'-5'-UTTTTTTTT-X- AATGATACGGCGACGGATCAAGT-Z-[TCO]3'	59 (91)	28471.6	48,76, 140	24+ ~50
C2b1	3'-tagagcatagcagaagaacgaac-5'-5'-UTTTTTTTT-X- AATGATACGGCGACGGATCAAGT-X- AATGATACGGCGACGGATCAAGT-Z-[TCO]3'	83(205)	64082.7	smear	-

Table 6.2: Overview of applied oligonucleotide for OSSeq optimization. A, C, G, T, U = 5'DMT-dA, dC, dG, dT, dU CE phosphoramidites; a, c, g, t = reverse phosphoramidites (dA-5'-CE, dC-5'-CE, dmF-dG-5'-CE and dT-5'-CE), X = S3 Linker, Z = 3'-amino-modifier CPG, Y = 5'-amino-modifier C6 (MMT). Bold are the stronger appearing bonds on gel. Not measured (-).

The final NHS-coupling with TCO-NHS-ester was done using standard NHS-condition from published procedure¹⁹⁵. TCO-NHS-ester (8 eq. per amine-modification) was added to each oligonucleotide to obtain TCO-primer oligonucleotides. However, the size of TCO when conjugated, only added 137.2 g/mol to the final structure per amino-modification, which made it very difficult to confirm successful conjugation with PAGE due to the small size difference. Instead, we decided to use the TCO-oligonucleotides in the assay and confirm the TCO-conjugation by successful conjugation with Tz-beads.

6.3.2 Optimization of OSSeqV2 (2.1)

Previously, GiWon Shin (GS) optimized the OSSeqV2 platform by exchanging the Taq polymerase (standard for PCR reactions) with a *Bst* DNA polymerase, which is a strand displacement enzyme, optimized for isothermal amplification¹⁹⁶. This polymerase is less sensitive to potential interruptions and therefore the extension happens more smoothly. He added SP1-blocker prior to the library extension which improved %On-target from 5% to 15%. Furthermore, the optimal bead-library ratio was determined to be 1/4, resulting in the current performance shown in **Table 6.1**. All optimization experiments done on OS-SeqV2 are illustrated in **Figure 6.9**. We tried to convert OS-Seq from APEX beads to Tz-tubes, however this was not successful, and the results are not shown here.

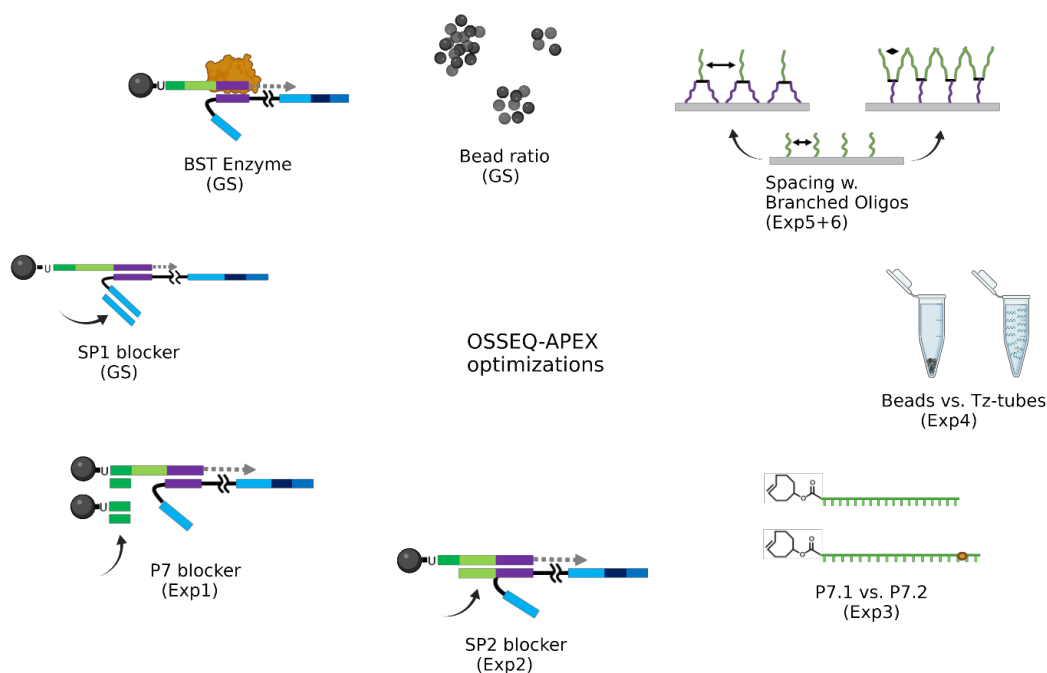


Figure 6.9: Experiments conducted to optimize OSSeqV2. Bead ratio, type of enzyme, SP1-blocker, P7-blocker, SP2-blocker, length of P7 primer, tubes vs. beads, and primer spacing using branched oligonucleotides. Experiments conducted by GS are stated in the figure.

6.3.2.1 Effect of P7- and SP2 Blocker in Library Capture Step

The OSSeqV2 method was applied as shown in **Figure 6.2** and the blockers were added to the library capture step prior to elongation. Library prepared (L1-L8) are shown in **Table 6.3** along with results of %On-target, %Ok and amount of library after USER (from qPCR). L1-L8 are merged in the table but were run as two experiments (L1-L4) and (L5-L8). L1-L4 all had SP1-blocker with varying parameters: Bst amount and P7-blocker (P7b). L5-L8 all had SP1-blocker and 0.5µl Bst enzyme with varying parameters being SP2-blocker (SP2b) and bead amount

(B:bead). It is seen from **Table 6.3** that L1-L8 all had %On-target around 25-30% and not affected much by either P7b or SP2b. However, looking at %Ok reads, P7b increase the percentage (~56%) while SP2b decrease the amount. An explanation of this could be that SP2b is too close to target binding area thereby affecting the quality of reads. Furthermore, it is observed that the amount of library after USER varies from L1-L4 to L5-L8 with a 50 fold increase. This is because larger fragments (~700bp) were used in L1-L4 while library size for L5-L8 was ~500bp. Larger fragments resulted in better %Ok reads, but less library is captured and elongated. We decided to aim for a fragment size between 500bp and 700bp for the next experiments. Best conditions from these experiments was SP1 blocker, 0.5µl Bst and P7 blocker.

Test	TCO-Primer	%On-target	%OK	Total amount a. USER (amol)
P7b	L1 (1ulBST)	31	37.7	4.0*
	L2 (1ulBST, P7b)	23	55.8	3.3*
	L3 (0.5ulBST)	31	31.0	5.9*
	L4 (0.5ulBST,P7b)	25	56.9	4.0*
SP2b	L5 (1xB)	26	27.8	138.1**
	L6 (1xB, SP2b)	26	16.3	193.7**
	L7 (0.5xB)	27	24.2	82.8**
	L8 (0.5xB, SP2b)	30	17.4	134.7**

Table 6.3. Effect of SP2 blocker and P7 blocker in OSSeqV2.

*Library size ~700bp, **Library size ~500bp.

6.3.2.2 Length of P7 Primer

The length of P7 primer was investigated by comparing two different lengths: P7.1 and P7.2, having extra 6nt, and a PS linkage in the 3'-end to avoid potential nuclease activity (**Table 6.2**). The two primers were tested with and without P7b and results shown in **Table 6.4**. It is seen that %On-target is slightly higher but similar for both primers around 25-31%. Moreover, the quality of reads was higher for P1.7, both with and without P7 blocker compared to P7.2 primer. Best conditions for this experiment were P7.1 primer with P7b.

TCO-Primer	%On-target	%OK
L1 (P7.1, P7b)	31	29.4
L2 (P7.1)	27	25.1
L3 (P7.2, P7b)	25	14.8
L4 (P7.2)	25	18.8

Table 6.4. Effect of P7.1 primer vs. P7.2 primer.

6.3.2.3 Effect of Spacing between Primers using Branched Oligonucleotides

For testing spacing between primer probes we tested P7.1, P7.2, Sb1, Sb2, Sb1.2, Sb1.3, Cb1, Cb2 with the conditions SP1 blocker, P7 blocker and 0.5µl Bst enzyme. The branched oligonucleotides were first conjugated to the beads, confirming the TCO-NHS-ester conjugation on amino-modified branched oligonucleotides. TCO-primers were added to Tz-beads and after 4 hours, no more oligonucleotide was removed from solution (**Figure 6.10**). The %conjugated TCO-primer was calculated from concentration difference between 0h and after 4h and results are seen in **Table 6.5**. Sb and Cb do not conjugate as well as P7.1 and P7.2, however they are also bulkier and occupy more space. Since no more primer was conjugated after 4 h, we expected beads were saturated with TCO-primer.

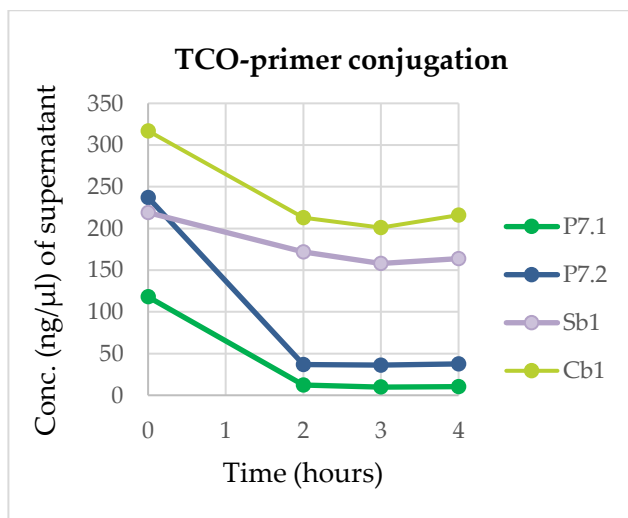


Figure 6.7. Conc. of TCO-primer in APEX solution during conjugation. P7.1, P7.2, Sb1 and Cb1 are shown as examples.









TCO-primer	%Conjugated
 P7.1	99.7
 P7.2	90.9
 Sb1	56.3
 S2b1	54.8
 Sb1.2	65.1
 Sb2.2	77.3
 Cb1	49.4
 C2b1	18.3

Table 6.5. Conjugated TCO-primer to beads, (% removed from solution).

After conjugation of branched TCO-primers, the primers were extended with targeting probes (OS42 pool). This process was monitored by measuring the concentration of removed oligo from solution before and after hybridisation. This showed unexpected results for the branched primers.

From the concentration of OS42 before and after, all the branched primers had hybridized between 0-10% (**Table S6.3**). These numbers were surprisingly low compared to P7.1 and P7.2 which hybridized 78% and 55% respectively. Therefore we decided to normalize the bead ratio, ensuring the beads contained 2.5pmol primer in all samples to be able to compare the individual primer performance in the next steps.

The procedure was continued with library capture using SP1b and P7b as optimal conditions. The extended library was cleaved with USER, showing a concentration around 4 amol (10^{-18} mol) for branched oligos. The library was amplified, quantified and sequenced. The results of %On-target reads and %Ok reads are shown in **Table 6.6**.

It is seen from the results that %On-target are in general low compared to previous experiments, also for control primers. This is because we accidentally used other adapters for the library prep which can self-ligate resulting in lower yield. However, it is also seen that all branched primers had 2-fold more %On-target reads (26-32%) compared with the two control sequence (15%). Moreover, the %Ok reads were worse for the branched primers compared to the controls, but among the branched primers, Cb1 had the highest quality similar to P7.2 primer. These findings are interesting, because the branched primers, especially Cb1, seems to have a good effect of number of reads captured in region of interest, and Cb1 has the same amount of OK reads as P7.2. This is also surprising since we assumed Sb structures would have the best effect, but from the results Cb1 has the highest %OK reads among the branched structures. It was also noted that C2b1 did not perform better than Cb1, concluding the amount of branches (N=2) did not improve the capture performance compared to N=1. However, these findings must be repeated with the right adapters, and small optimizations in the assay to confirm the results.

TCO-Primer	%On-target	%OK
L1: P7.1	14	11.6
L2: P7.2	15	7.0
L3: Sb1	28	3.5
L4: S2b1	28	4.5
L5: Sb1.2	27	4.0
L6: Sb2.2	32	3.6
L7: Cb1	26	6.1
L8: C2b1	29	3.7

Table 6.6. Extracted sequencing data evaluating the performance of branched primers in OS-SeqV2.

6.3.3 Summary of OS-SeqV2 optimizations

In summary, we investigated the effect of P7-blocker, SP2-blocker, primer length and primer space to improve capture performance of current Oligonucleotide-Selective Sequencing method (OS-SeqV2). We evaluated the performance in terms of %On-target reads and %Ok reads from sequencing data. P7-blocker improved %On-target from 20% to 34% and %Ok reads increased from 10% to 32%. SP2-blocker or a longer primer (P7.2) did not improve the assay. Primer spacing were investigated using synthesized *spacing* primers (Sb1, S2b1, Sb1.2, Sb2.2) and *crowding* primers (Cb1 and C2b1). Branching primers seem to improve %On-target 2-fold compared to control primers which is promising. Cb1 had the best %Ok reads of the branched primers similar to P7.2 control sequence. The amount of library captured and elongated on the beads are less (~5 amol) compared to P7.1 primer (~100 amol), and therefore not suitable in potential PCR-free assay. The optimal loading concentration for e.g. NextSeq platform is 1.5pM but can work in fmol scale¹⁹⁷. Therefore, the assay needs to be further optimized to reach a PCR-free stage. However, branched primers seem to have a good effect on %On-target reads, compared with controls and is something we want to confirm in the future. The performances of OS-SeqV1, OS-SeqV2 and newest optimization OS-SeqV2.1 are shown below.

Condition	V1 (Run633, L006942) ¹⁸⁸	V2 (SP1b,1/4 beads)*	V2.1 (P7.1, SP1b, P7)**
%On-target	82%	20%	34%
On target read pair	8,325,559	110,218	949,338
Ok reads	7,146,544	20,978	609,852
%Ok	42.92%	9.52%	32.2%

Table 6.7: Summary table of OSSeq performance of first versus second generation of OS-Seq.

*Numbers of V2 obtain by GiWon Shin (GS) and reprinted with permission. **Potential PCR-free.

6.3.4 Future work

In the future, we want to confirm the mass of branched oligonucleotides with MS, and purify to validate the results. Furthermore, we want to repeat the primer spacing experiment with the right adapters and few optimizations on the assay to confirm performance of branched primers improving %On-target with 100% compared to control sequences. When the best conditions have been established, we want to test the optimized OS-SeqV2.1 in cancer plasma samples to determine selectivity compared to whole-genome sequencing.

6.4 Experimental

6.4.1 General Chemical Experimental Methods

Chapter 5, section 5.4.1.

6.4.1.1 Synthesis of Branched Linker (S3)

S3 is described in Chapter 5, section 5.4.2.

6.4.2 General Methods for Molecular Assays

All reagents and enzymes were purchased from either Roche, New England Biolabs, Invitrogen, Cube Biotech or PolyAn. Genomic DNA sample used in this study was NA12822 (428 ng/ul) from Coriell Institute. Buffers were freshly prepared using lab procedure. PCR amplification (KAPA HiFi HotStart kit from Roche with Illumina primers 10X mix (P7 and P5)) was performed on Applied Biosystems Veriti 96 well Thermal Cycler. RT-qPCR quantification was performed using Applied Biosystems StepOnePlus Real-Time PCR-systems with RT-qPCR-kit from Roche (KAPA SYBR mastermix, KAPA primer mix, ROX High, Standards 1-6: 20, 2, 0.2, 0.02, 0.002, 0.0002 pM). Fluorescence quantification of dsDNA or ssDNA was measured using Qubit4 Fluorometer from Invitrogen with Qubit 1x dsDNA BR Assay kit and Qubit ssDNA Assay kit. Size characterization of library was made using E-Gel EX Agarose Gels, 2% from Invitrogen, with 100bp ladder from NewEngland Biolabs. Gels were run for 9min on E-Gel Power Snap Electrophoresis device, and visualized with E-Gel Power Snap Camera both from Invitrogen. Sequencing was performed on iSeq 100 i1 from Illumina with Cartridge and Flow Cell. PhiX (50pM) was used as reference genome for alignment.

6.4.3 Oligonucleotide Synthesis

All branched oligonucleotides were synthesized on Biosset DNA/RNA Synthesizer in 0.1μmol scale using universal support (CUTAG CPG 1000 Å from Sigma Aldrich) for Sb and 3'-Amino-Modifier C7 CPG 1000 (Z), Glen Reaserch for Cb. Reagents applied for automated solid-phase DNA synthesis were dry MeCN, TCA Deblock, DCI Activator 0.25M, CapA, CapB, Oxidizer 0.02M from Sigma Aldrich. Phosphoramidites dA(bz), dC(bz), dG(ib), dT and dU (Sigma Aldrich and Glen Research) were dissolved with dry MeCN in 0.1M. Furthermore, reverse phosphoramidites (dA-5'-CE, dC-5'-CE, dmF-dG-5'-CE and dT-5'-CE from Glen Research) were

also prepared for Cb in dry MeCN (0.1M). Branched Linker (X: S3) were coupled manually (12 mg in 400µl MeCN, 600 µl DCI Activator over 20 min). MMT-protected 5'-Amino-Modified phosphoramidite (Y) was dissolved in dry MeCN (0.1M) and applied for all Sb oligonucleotides. All oligonucleotides were synthesized in DMT-off mode with extra deprotection provided on the last base to insure complete removal of MMT from the amines. The oligonucleotides were dried with N₂ and put into a 5mL vial, treated with ammonia, 55°C, o/n to remove solid-support, protecting groups, Fmoc and MMT amine protecting groups. The oligonucleotides were cooled down, filtered into Eppendorf tubes, washed with EtOH:MQ (1:1), and solvent removed on vacuum centrifuge to dryness. The oligonucleotides were dissolved in 1000µL MQ water and quantified with nanodrop. The length of the oligonucleotides were identified using Novex™ TBE-Urea Gels 10% and USER enzyme. Oligonucleotides were not purified.

Name	Sequences 5'→3'	MW [g/mol]	Length [nt]	Length obs. [nt]
Sb1	5'Y-AATGATACGGCGACGGATCAAGT-X- TTTTTTTTTUC AAGCAGAAGACGGC ATACGAGAT-3'	25285.6	59 (80)	~140
S2b1	5'Y-AATGATACGGCGACGGATCAAGT-X- AATGATACGGCGACGGATCAAGT-X- TTTTTTTTTUC AAGCAGAAGACGGC ATACGAGAT-3'	54584.9	83 (172)	Smear
Sb1.2	5'Y-CGACGGATCAAGT-X- TTTTTTTTTUC AAGCAGAAGACGGC ATACGAGAT-3'	19009.4	49 (63)	34, 50, 78
Sb2.2	5'Y-CGACGGATCAAGT-X- TTTTTTTTTUC AAGCAGAAGACGGC ATACGAGATCGGTCT-3'	20854.6	55 (69)	40, 56, 84
Cb1	3'-tagagcatagcgagaagacgaac-5'-5'-UTTTTTTTT-X- AATGATACGGCGACGGATCAAGT-Z-3'	28471.6	59 (91)	48, 76, ~140
C2b1	3'-tagagcatagcgagaagacgaac-5'-5'-UTTTTTTTT-X- AATGATACGGCGACGGATCAAGT-X- AATGATACGGCGACGGATCAAGT-Z-3'	64082.7	83 (205)	Smear

Table S6.1: Oligonucleotide Synthesis and Characterization. Y: 5'-Amino-Modifier-C6 (MMT), X: S3 Branched Linker, Z: 3'-Amino-Modifier-C7 CPG (Fmoc, FW: 209.18g/mol), U=dU, lower case: revers DNA phosphoramidites, upper case: DNA phosphoramidites.

6.4.4 Gel Characterization of Branched Oligonucleotides

6.4.4.1 Sample Preparation

Dry oligonucleotides were dissolved to stock concentration 1000 ng/µl. Dilution (1:100) was made to reach conc. 10 ng/µl, and 5µl of each dilution was mixed with 5µl loading dye (NBE) and denatured before loaded on gel (95°C, 3min and on ice). 10bp ladder applied from NBE, and prepared in 1:10 dilution in MQ with addition of 5µl loading dye to reach 10µl.

6.4.4.2 Gel Preparation

Oligonucleotides were characterized using Novex™ TBE-Urea Gels 10% from Invitrogen. The gel was rinsed with demi-water (comb and tape removed) and pre run at 160V for 30 min with 1xTBE as running buffer solution. Running buffer was pipetted into each well twice (1000µl) to remove urea and samples loaded (10µl, 50 ng in each well) and gel run at 180V. When finished, the gel was released and stained with SYBRGold in 1xTBE (5min). Gel was visualized on Gel Dox XR+ Systems from Bio-Rad with UV light.

6.4.4.3 Treatment with USER

For treatment with USER, the thermomixer was set to 37°C, then dilutions was made and enzyme mix made. USER mix was made by adding 1U/µl USER enzyme (11µl) to 10X CutSmart buffer (22µl) and 187µl water (4 rxn of 50µl). 1:10 dilutions was made from stock conc. of Sb1, Sb2.1, Sb2.2 and Cb1. 5µl oligonucleotide dilutions was transferred to tubes with 50µl USER enzyme mix and left for react on thermomixer o/n (18h) at 1400 rpm. 5µl oligonucleotide+USER mix was transferred to new tubes with addition of 5µl loading dye, and loaded on gel as described above.

6.4.5 NHS coupling

Each oligonucleotide (4nmol) was taken from a stock (1000 ng/µl) and transferred to a new Eppendorf tube and evaporated to almost dryness. Bicarbonate buffer (108µl, 0.1M) was added, followed by addition of DMSO (32.8µl/36.8µl). A freshly prepared solution of TCO-NHS ester (BroadPharm) in DMSO (1mg/300µl) was added to each oligonucleotide solution with 8 eq. per reactive amino site: Sb1/Sb2.1/Sb2.2 (16eq.); S2b1 (32eq.), Cb1/C2b1 (8eq.). The reaction is mixed by vortex and put on invert rotation o/n at rt to react. After 24h, the tubes were spin down, added 1ml cold acetone to precipitate. The tubes were put in freezer overnight to complete precipitation of conjugated oligonucleotides. The pellets were centrifuged 10 min, supernatant removed and washed with additional 1ml cold acetone followed by centrifuge. Washing step was repeated 3 times. In the final wash, excess acetone was removed using vacuum concentrator and the dry pellet dissolved in MQ to a final conc. of 100µM for each branched TCO-oligonucleotide. Due to the large size difference between oligonucleotide (ranging from 19009.4-64082.7 g/mol) and conjugated TCO moiety (~150g/mol) it was not possible to see the difference in gel. MALDI-TOF was not available at the site. Therefore, we confirmed the conjugation by using the TCO-primers in the assay.

6.4.6 OSSEQ-APEX Method

6.4.6.1 Adaptor-preparation

Adapters design is shown in **Figure 6.2G** and sequences applied are shown in **Table S6.2**, with top-part containing P5 (blue), i5 (red) and SP1 (black) and phosphorthioate in 3'-end. The universal bottom was complementary to SP1 having dideoxy-cytosine (ddC) in 3'-end. The oligonucleotides were purchased from IDT and prepared in stocks of 100µM. The adapter pairs were annealed by mixing respective adapter top (7.5µl, 100µM) with universal bottom (7.5µl, 100µM), with 15µl Nuclease Free Duplex buffer (IDT) to a final volume and conc. on 30µl, 15µM. The mixture were run on PCR thermocycler: 95°C for 5 min, then ramp down to 20°C with rate of 1.5°C/min and stored in -20°C prior to use. Final adapter pairs were named D501-D508.

Oligo ID	Sequence 5'→3' (P5-i5-SP1)
Ad_top_OSSEQ_D501	AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTCCCT ACACGACGCTCTTCCGATC*T
Ad_top_OSSEQ_D502	AATGATACGGCGACCACCGAGATCTACACATAGAGGCACACTCTTCCC TACACGACGCTCTTCCGATC*T
Ad_top_OSSEQ_D503	AATGATACGGCGACCACCGAGATCTACACCTATCCTACACTCTTCCCT ACACGACGCTCTTCCGATC*T
Ad_top_OSSEQ_D504	AATGATACGGCGACCACCGAGATCTACACGGCTCTGAACACTCTTCCCT ACACGACGCTCTTCCGATC*T
Ad_top_OSSEQ_D505	AATGATACGGCGACCACCGAGATCTACACAGGCGAAGACACTCTTCCC TACACGACGCTCTTCCGATC*T
Ad_top_OSSEQ_D506	AATGATACGGCGACCACCGAGATCTACACTAATCTTAACACTCTTCCCT ACACGACGCTCTTCCGATC*T
Ad_top_OSSEQ_D507	AATGATACGGCGACCACCGAGATCTACACAGGACGTACACTCTTCCC TACACGACGCTCTTCCGATC*T
Ad_top_OSSEQ_D508	AATGATACGGCGACCACCGAGATCTACACGTACTGACACTCTTCCCT ACACGACGCTCTTCCGATC*T
Ad_bot_OSSEQ_i5_universal_ ddC	/5Phos/GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3ddC/

Table S6.2: OS-Seq adapter sequences, top and universal bottom. *Phosphothioate bond; Phos: Phosphate; ddC: dideoxy cytosine; P5 (blue), i5 (red), SP1 (black).

6.4.6.2 Library Preparation

The library was prepared from a genomic DNA sample (NA12822, 428 ng/µl, Coriell Institute). First a dilution of the sample (132µl, 33.33 ng/µl) was enzymatically fragmented by adding 22µl diluted conditioning solution (3.1µl in 96.9µl water), 22µl 10X reaction buffer, and 44µl fragmentation enzyme to the diluted sample and put on ice. The reaction mixture was mixed well by pipet and divided into 4 PCR-tubes (each 50µl), incubated for 17 min at 37°C in a thermocycler and immediately put on ice. A reaction mixture of End repair & A-tailing buffer (35µl) and enzyme mix (15µl) was made and added to each tube (10µl) to add A-tails to the 5'end of the fragments. Tubes were incubated at 60°C for 30 min. Next, the adapters (15µM, D501-D504 or

D505-D508) were added (5 μ l) to each tube (65 μ l in total) and a ligation mixture was made by adding water (22 μ l), ligation buffer (132 μ l) and DNA ligase (44 μ l) and 45 μ l was added to each tube and incubated for 30 min at 20°C. The library was purified with magnetic OMEGA beads (0.8x (88 μ l)), incubated for 5 min at rt., washed twice with 200 μ l 80% EtOH, dried for 3 min and eluted with 54.5 μ l 10mM Tris. To characterize library size, 2% agarose gel was made by loading samples (1:10 dilution) and 100bp ladder (1:20 dilution) to the gel and run E-Gel Power Snap Invitrogen (Thermo Fisher Scientific) for 8 min and visualized by UV. The library was kept at -20°C until used for library capture.

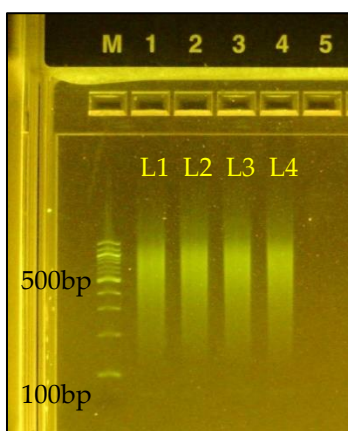


Figure S6.1: Example of Library prep (L1-L4) in 2% Agarose, with size ~500bp.

6.4.6.3 Primer conjugation

An aliquot of primer stock (TCO-P7 or TCO-P7.2, 10 μ l, 100 pmol/ μ l) was diluted with 94 μ l TrisTween buffer and 2 μ l was saved for quantification (1:10, 0h) for each primer. Magnetic tetrazine beads (APEX) was transferred to new PCR-tubes (10 μ l in each) for each primer and washed twice with 100 μ l TrisTween buffer. The remaining primer dilution (2x51 μ l) was transferred to the washed beads (51 μ) and incubated for 2h at rt. by inverting rotation. The supernatant was kept for quantification measurement by Qubit (ssDNA kit) and primer conjugated beads were stored in APEX-storage buffer at -20°C until next step.

6.4.6.4 Primer extension with targeting oligonucleotide pool

The oligonucleotide pool (OS42) was available in the Ji Research Lab, and was similar to the one applied and described in the paper from Shin et al. in supplementary Information (Additional file 2 and 3)¹⁹². However, OS42 did not contain probes targeting the *MSH3* gene. The OS42 oligonucleotide pool contained in total 6205 different primer probes targeting 84 different genes (target size 411.2 kb.). The oligo pool library (OS42) was added to the beads for hybridization and

extension. The stored APEX were washed twice with 100µl TrisTween buffer and once with 100µl hybridization buffer 1 (HT1). A dilution of oligo pool was made by adding 4 x hybridization buffer (19.3µl) and water (30.3µl) to 2µM OS42 oligo stock solution (27.5µl), and 70µl of the diluted oligo solution was added to the APEX, mixed on thermomixer 96°C, 5min, 1400rpm, and then cooled to 40°C while on mixer (1400rpm). 2XQ5U mix (polymerase) was diluted with water (50:50), hot started on thermocycler 95°C in 3min. The supernatant from beads were saved for quantification measurement, beads washed twice with hybridization buffer 2 (HT2) and Q5U mix was added (70µl) and mixed at 74°C, 10min at 800rpm, then 2min with 1400rpm. The supernatant was saved for quantification and formamide buffer (70µl) was added twice and incubated at 96°C for 5min at 1400rpm to denature any hybridization to the extended primer probes (both supernatants were kept). Post denaturation, the beads were washed with TrisTween buffer (70µl), incubated at 96°C for 5min at 1400rpm (supernatant kept), and finally washed with 100µl HT1. The extended primer probe beads were split into 4 aliquots (25µl each) and used directly for library capture. Supernatants measured by Qubit (ssDNA kit). For experiment with branched primers, hybridization did not occur completely, so the libraries were normalized to the primer with lowest amount of extended primer probe (Cb1 with 2.5pmol). All primers were normalized to have 2.5pmol extended primer on beads. The two cases with negative values were calculated as 100µl.

Primer probe	Conc. after hybridization [ng/µl]	Total amount in solution [pmol]	Hybridized primer probe [pmol]	Beads transferred for 2.5pmol [µl]
OS42 2µM	36.1	81		
P7.1	7.73	17	63.6	4
P7.2	16.1	36	44.8	5
Sb1	37.3	84	-2.7	100
S2b1	37.6	84	-3.4	100
Sb1.2	32.6	73	7.8	31
Sb2.2	31.7	71	9.9	25
Cb1	35	78	2.5	100
C2b1	34.7	78	3.1	79

Table S6.3: Examples of Probe hybridization (OS42) to branched primers. Conc. of OS42 in solution after hybridization [ng/µl]; Total amount of OS42 in solution [pmol]; Hybridized primer probe (101nt) [pmol] and Beads transferred for 2.5pmol in each tube [µl].

6.4.6.5 Library Capture

The prepared library was added specific blockers (P7b, SP2b, 1µl), water (1µl) and 4X Hyb buffer (17.5µl all wells) depending on the experiment and incubated for 15min at 95°C and immediately put on ice. HT1 from beads were removed and the library containing blockers was added to the beads (70µl) and incubated 5min at 96°C, 1400rpm, then cooled to 65°C while 1400rpm and left

for >20h on shaker 65°C, 1400 rpm. The polymerase mixture (Bst) was made by adding 10X isothermal amplification buffer (30.8µl), dNTP 10mM each (6.2µl), Bst 2.0 warmstart (2.2µl) in water (268.8µl). The supernatant from after hybridization was kept (dsDNA kit), beads washed twice with 100µl HT2 and the Bst reaction mixture was added (70µl) to the beads, incubated for 10 min at 65°C with 800rpm, then 2min at 65°C with 1400rpm (supernatant saved after this step, dsDNA kit). To denature the library from extended probes, formamide buffer (70µl) was added to the beads twice, incubated at 96°C for 5min at 1400rpm, and washed once with TrisTween buffer (70µl), 96°C for 5min at 1400rpm. Supernatant from first denaturation was kept for quantification with Qubit (ssDNA), and beads used directly in next step.

6.4.6.6 Cleavage from APEX (USER)

The beads were washed with 100µl 1XCutSmart buffer and added 70µl of the USER enzyme mixture (1U/u USER enzyme 15.4µl, 10XCutSmart buffer (30.8µl), water (261.8µl)) in each tube, and incubated for 19h at 37°C with 1400rpm. The cleaved library (supernatant) was transferred to new tubes and the beads was washed with 50µl TrisTween buffer and stored in APEX-storage buffer (70µl) at -20°C. The library was purified with OMEGA beads (1.8x (126µl)), incubated for 15min at rt., washed twice with 200µl 80% EtOH, dried for 3min and eluted with 22µl 10mM Tris. 2µl of the cleaned library was diluted with water and used to determine concentration with Qubit and qPCR. The library was kept at -20°C until ready to pool for sequencing.

6.4.6.7 Quantification and Amplification

qPCR after USER cleavage was performed by making the qPCR reaction mixture (11 samples: KAPA SYBR master mix (110µl), water (72.6µl), KAPA primer mix (22µl), ROX High (22µl)) which was added to a qPCR 96 well plate. Standards 1-6 (20, 2, 0.2, 0.02, 0.002, 0.0002 pM) and samples were diluted 1:20 with qPCR reaction mixture and mixed well, sealed with sticky plastic cover, centrifuged to avoid bubbles, and run on Real-Time qPCR, Applied Biosystem. qPCR settings were 95°C for 5min, 40 cycles of (95°C in 30s + 60°C in 90s), 72°C for 7min. 2% agarose gel was performed on the qPCR product after USER to control size. Results from qPCR are shown below with example of qPCR curves and calculated yield after USER for branched primer (2.5pmol).

PCR were performed on the cleaned library by adding 5µl Illumina primer 10X mix and 25µl Kapa HiFi Hot Start to each library and mixed well (total volume 50µl). The PCR machine was set to 98°C for 5min, 12 cycles of (98°C in 15s + 60°C in 30s + 72°C in 30s), then 1min of 72°C and hold on 4°C. After PCR the amplified captured library was purified with OMEGA beads (1x vol.),

incubated for 5min at rt., washed twice with 200µl 80% EtOH, dried for 3min and eluted with 30µl 10mM Tris.

After PCR-amplification another qPCR and 2% agarose gel were made to determine concentration and library size. 2µl of the amplified captured library was diluted 1:10000 in water and used for qPCR and gel in the same settings and conditions as described above.

Primer	pM, 452bp std.	Original conc. (pM)	Total conc. after USER [amol]	Total conc. after PCR [amol]
P7.1	0,0181	0,1812	3,9859	4835
P7.2	0,0224	0,2237	4,9203	10151
Sb1	0,0124	0,1237	2,7212	3827
Sb2	0,0077	0,0770	1,6948	1895
Sb1.2	0,0087	0,0867	1,9083	1559
Sb1.3	0,0079	0,0795	1,7481	2299
Cb1	0,0284	0,2840	6,2479	7129
Cb2	0,0112	0,1118	2,4593	4779

Figure S6.4: Yield of libraries from branched primers (qPCR) after USER and yield after PCR.

6.4.6.8 Sequencing

The iSeq 100 i1 reagent kit from Illumina (cartridge) was taken out of the freezer and put in +4°C ~1.5-2 days before sequencing. The library were pooled to obtain a final volume of 40µl with a concentration of 1nM. The final sequencing reaction mixture was made by adding 1µl of the pooled library (1nM) and 2µl 50pM PhiX (9%) to 19µl 10mM Tris. The mixture was mixed well and added to the cartridge that has been shaken and flow cell added. The cartridge with flow cell and loaded sample (20µl) were put into iSeq 100 sequencer, and script made to prepared for sequencing. The sequencing took around 18 h, for the run to finish. Data was extracted using the OS-Seq pipeline, number of reads, on-target reads and OK reads were extracted. These were then evaluated and performance compered as shown in results.

7 Conclusion

In the thesis, three strategies to optimize current FLEET platform established in Astakhova group, were investigated. First, by using calculation from the Peyrard-Bishop model, insertions of LNA in high-affinity capture sequences targeting oncogenes, could be designed and their T_m predicted. Selected probes were synthesized, and had their T_m measured, resulting in an accuracy of 1°C. This indicated the PB model could indeed be applied for future design of ultra-specific probes.

Secondly, we aimed to implement the FLEET assay in microfluidic PMMA chips using TC-tagged capture probes immobilized with UV light. For this purpose, we fabricated microfluidic PMMA chips and estimated LOD to be 6.4pM – 32pM using a spectrofluorometer. The assay was tested with BRAF-probes, designed with the PB model, showing detectable signal. However, even after repeated attempts of improvement, the signal remained low and inconsistent. Therefore, further optimizations are needed to reach a satisfactory result.

To expand the FLEET assay, we designed specific hybridization probed to detect SNPs in miRNA-128-2-3p. Again, we applied the PB model to design LNA enriched capture probes with high specificity for miRNA128-2-3p and its mutants. We showed high discrimination between match and mismatch probes with ΔT_m range between 6.6-25.5°C. Using the assay, we successfully quantified levels of miRNA128-2-3p and its mutated variants in plasma samples from colitis and colorectal cancer patients. A significantly higher level of wild type miRNA128-2-3p in colitis and colorectal cancer than in healthy was detected. Moreover, a generally higher amount of mutated miRNA128-2-3p in colitis and CRC than in healthy controls, except for mutations in the 3'-3 position was observed.

Related to production of probes, the thesis presents a strategy to make pools of miRNA using TONS by integrating synthesized 2'/3'OAc uridine phosphoramidites (normal and reverse). In combination these phosphoramidites, act as cleavable site during the solid-phase synthesis. We successfully made two, three and four consecutive microRNAs (miR129-1-5p, miR31, miR206, miR27b-3p) using linear pool strategy. Furthermore, we investigated the effect of branching oligonucleotides aiming for increase in yield. However, the overall yields were not improved compared to linear approach. Our cleavable cite approach is a promising way to make microRNA pools and can contribute to the synthesis of RNA therapeutics in the future.

Targeted sequencing has great importance in finding new variants of nucleic acid biomarkers. In this thesis, we try to optimize a targeting sequencing method, *Oligonucleotide-Selective Sequencing*, established in the Ji Research lab. Specifically, we investigate the effect of primer spacing using synthesized branched oligonucleotide as primers, with the hope of improving the yield and sensitivity. The branching primers were found to improve %On-target by 2-fold compared to control primers which is promising. The crowding branched oligo had the best performance of the branched primers. Unfortunately, the yield is decreased with branched primers, and therefore not suitable for potential PCR-free application. However, these findings need to be confirmed in near future.

8 Future Perspective

Implementing the FLEET assay in microfluidic chips in the course of making a PoC platform is an exciting and challenging interdisciplinary task. However, many optimizations steps are needed to reach a final product. The FLEET platform still has potential to be a PoC device. A future strategy could be to ensure the chemistry is working on established platforms similar to microarrays before moving to cheaper options. Therefore, we would like to test the assay on another material than PMMA in near future, such as glass slides functionalized with aldehyde. Furthermore, we should ensure optimized conditions in every step using established detection for microarrays, such every step can be followed.

For the OS-Seq method, we want to confirm the mass of branched oligonucleotides with MS, and purify to validate the results. Furthermore, we want to repeat the primer spacing experiment with the right adapters and few optimizations on the assay to confirm performance of branched primers improving %On-target with 100% compared to control sequences. When the best conditions have been established, we will test the optimized OS-SeqV2.1 with colon cancer plasma samples to determine selectivity and compare with whole-genome sequencing.

References

- (1) *Cancer*. <https://www.who.int/news-room/fact-sheets/detail/cancer> (accessed 2023-01-23).
- (2) Knight, S. B.; Crosbie, P. A.; Balata, H.; Chudziak, J.; Hussell, T.; Dive, C. Progress and Prospects of Early Detection in Lung Cancer. *Open Biol.* **2017**, *7* (9). <https://doi.org/10.1098/RSOB.170070>.
- (3) Cui, M.; Wang, H.; Yao, X.; Zhang, D.; Xie, Y.; Cui, R.; Zhang, X. Circulating MicroRNAs in Cancer: Potential and Challenge. *Front. Genet.* **2019**, *10*, 626. <https://doi.org/10.3389/FGENE.2019.00626>.
- (4) Condrat, C. E.; Thompson, D. C.; Barbu, M. G.; Bugnar, O. L.; Boboc, A.; Cretoiu, D.; Suciu, N.; Cretoiu, S. M.; Voinea, S. C. MiRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. *Cells* **2020**, *9* (2). <https://doi.org/10.3390/CELLS9020276>.
- (5) Gambardella, V.; Tarazona, N.; Cejalvo, J. M.; Lombardi, P.; Huerta, M.; Roselló, S.; Fleitas, T.; Roda, D.; Cervantes, A. Personalized Medicine: Recent Progress in Cancer Therapy. *Cancers (Basel)*. **2020**, *12* (4). <https://doi.org/10.3390/CANCERS12041009>.
- (6) Lynch, T. J.; Bell, D. W.; Sordella, R.; Gurubhagavatula, S.; Okimoto, R. A.; Brannigan, B. W.; Harris, P. L.; Haserlat, S. M.; Supko, J. G.; Haluska, F. G.; Louis, D. N.; Christiani, D. C.; Settleman, J.; Haber, D. A. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non-Small-Cell Lung Cancer to Gefitinib. *N. Engl. J. Med.* **2004**, *350* (21), 2129–2139. <https://doi.org/10.1056/NEJM0A040938>.
- (7) Paez, J. G.; Jänne, P. A.; Lee, J. C.; Tracy, S.; Greulich, H.; Gabriel, S.; Herman, P.; Kaye, F. J.; Lindeman, N.; Boggon, T. J.; Naoki, K.; Sasaki, H.; Fujii, Y.; Eck, M. J.; Sellers, W. R.; Johnson, B. E.; Meyerson, M. EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy. *Science* **2004**, *304* (5676), 1497–1500. <https://doi.org/10.1126/SCIENCE.1099314>.
- (8) Diaz, L. A.; Williams, R. T.; Wu, J.; Kinde, I.; Hecht, J. R.; Berlin, J.; Allen, B.; Bozic, I.; Reiter, J. G.; Nowak, M. A.; Kinzler, K. W.; Oliner, K. S.; Vogelstein, B. The Molecular Evolution of Acquired Resistance to Targeted EGFR Blockade in Colorectal Cancers. *Nature* **2012**, *486* (7404), 537–540. <https://doi.org/10.1038/nature11219>.
- (9) Yoneda, K.; Imanishi, N.; Ichiki, Y.; Tanaka, F. Treatment of Non-Small Cell Lung Cancer with EGFR-Mutations. *J. UOEH* **2019**, *41* (2), 153–163. <https://doi.org/10.7888/JUOEH.41.153>.
- (10) Antolín, S.; Calvo, L.; Blanco-Calvo, M.; Santiago, M. P.; Lorenzo-Patiño, M. J.; Haz-Conde, M.; Santamarina, I.; Figueroa, A.; Antón-Aparicio, L. M.; Valladares-Ayerbes, M. Circulating MiR-200c and MiR-141 and Outcomes in Patients with Breast Cancer. *BMC Cancer* **2015**, *15* (1). <https://doi.org/10.1186/S12885-015-1238-5>.

- (11) Zanutto, S.; Pizzamiglio, S.; Ghilotti, M.; Bertan, C.; Ravagnani, F.; Perrone, F.; Leo, E.; Pilotti, S.; Verderio, P.; Gariboldi, M.; Pierotti, M. A. Circulating MiR-378 in Plasma: A Reliable, Haemolysis-Independent Biomarker for Colorectal Cancer. *Br. J. Cancer* **2014**, *110* (4), 1001. <https://doi.org/10.1038/BJC.2013.819>.
- (12) Zhang, J.; Zhang, K.; Bi, M.; Jiao, X.; Zhang, D.; Dong, Q. Circulating MicroRNA Expressions in Colorectal Cancer as Predictors of Response to Chemotherapy. *Anticancer. Drugs* **2014**, *25* (3), 346–352. <https://doi.org/10.1097/CAD.0000000000000049>.
- (13) Zhao, Y.; Song, Y.; Yao, L.; Song, G.; Teng, C. Circulating MicroRNAs: Promising Biomarkers Involved in Several Cancers and Other Diseases. *DNA Cell Biol.* **2017**, *36* (2), 77–94. <https://doi.org/10.1089/DNA.2016.3426>.
- (14) Diehl, F.; Schmidt, K.; Choti, M. A.; Romans, K.; Goodman, S.; Li, M.; Thornton, K.; Agrawal, N.; Sokoll, L.; Szabo, S. A.; Kinzler, K. W.; Vogelstein, B.; Diaz, L. A. Circulating Mutant DNA to Assess Tumor Dynamics. *Nat. Med.* **2008**, *14* (9), 985–990. <https://doi.org/10.1038/NM.1789>.
- (15) Wang, D.; Farhana, A. Biochemistry, RNA Structure. *StatPearls* **2022**.
- (16) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nat.* 1953 1714356 **1953**, *171* (4356), 737–738. <https://doi.org/10.1038/171737a0>.
- (17) Vitz, E.; Bruist, M. F.; Smith, W. L.; Mell, G. A Simple Demonstration of How Intermolecular Forces Make DNA Helical Tested Demonstrations A Simple Demonstration of How Intermolecular Forces Make DNA Helical. *J. Chem. Educ. Abstr. Test. Demonstr. Classr. JChemEd.chem.wisc.edu* • **1998**, *75* (1).
- (18) Nikitin, M. P. Non-Complementary Strand Commutation as a Fundamental Alternative for Information Processing by DNA and Gene Regulation. *Nat. Chem.* 2023 151 **2023**, *15* (1), 70–82. <https://doi.org/10.1038/s41557-022-01111-y>.
- (19) Chatterjee, N.; Walker, G. C. Mechanisms of DNA Damage, Repair and Mutagenesis. *Environ. Mol. Mutagen.* **2017**, *58* (5), 235. <https://doi.org/10.1002/EM.22087>.
- (20) Zemora, G.; Waldsich, C. RNA Folding in Living Cells. *RNA Biol.* **2010**, *7* (6), 634. <https://doi.org/10.4161/RNA.7.6.13554>.
- (21) Sato, K.; Akiyama, M.; Sakakibara, Y. RNA Secondary Structure Prediction Using Deep Learning with Thermodynamic Integration. *Nat. Commun.* 2021 121 **2021**, *12* (1), 1–9. <https://doi.org/10.1038/s41467-021-21194-4>.
- (22) *The Genetics of Cancer* - NCI. <https://www.cancer.gov/about-cancer/causes-prevention/genetics> (accessed 2023-01-23).
- (23) *Lung Cancer Risk Factors | Smoking & Lung Cancer*. <https://www.cancer.org/cancer/lung-cancer/causes-risks-prevention/risk-factors.html> (accessed 2023-01-23).
- (24) Wang, Y.; Ru, J.; Meng, X.; Song, J.; Jiang, Q.; Li, S.; Jiang, J.; Li, Y. Role of SNPs in the Biogenesis of Mature MiRNAs. *Biomed Res. Int.* **2021**, 2021.

- <https://doi.org/10.1155/2021/2403418>.
- (25) Chevallier, M.; Borgeaud, M.; Addeo, A.; Friedlaender, A. Oncogenic Driver Mutations in Non-Small Cell Lung Cancer: Past, Present and Future. *World J. Clin. Oncol.* **2021**, *12* (4), 217. <https://doi.org/10.5306/WJCO.V12.I4.217>.
 - (26) Crosby, D.; Bhatia, S.; Brindle, K. M.; Coussens, L. M.; Dive, C.; Emberton, M.; Esener, S.; Fitzgerald, R. C.; Gambhir, S. S.; Kuhn, P.; Rebbeck, T. R.; Balasubramanian, S. Early Detection of Cancer. *Science* (80-.). **2022**, 375 (6586). <https://doi.org/10.1126/SCIENCE.AAY9040/ASSET/0A3087E0-148F-45C4-B9D9-5558E774A080/ASSETS/IMAGES/LARGE/SCIENCE.AAY9040-F6.JPG>.
 - (27) Bettegowda, C.; Sausen, M.; Leary, R. J.; Kinde, I.; Wang, Y.; Agrawal, N.; Bartlett, B. R.; Wang, H.; Luber, B.; Alani, R. M.; Antonarakis, E. S.; Azad, N. S.; Bardelli, A.; Brem, H.; Cameron, J. L.; Lee, C. C.; Fecher, L. A.; Gallia, G. L.; Gibbs, P.; Le, D.; Giuntoli, R. L.; Goggins, M.; Hogarty, M. D.; Holdhoff, M.; Hong, S. M.; Jiao, Y.; Juhl, H. H.; Kim, J. J.; Siravegna, G.; Laheru, D. A.; Lauricella, C.; Lim, M.; Lipson, E. J.; Marie, S. K. N.; Netto, G. J.; Oliner, K. S.; Olivi, A.; Olsson, L.; Riggins, G. J.; Sartore-Bianchi, A.; Schmidt, K.; Shih, I. M.; Oba-Shinjo, S. M.; Siena, S.; Theodorescu, D.; Tie, J.; Harkins, T. T.; Veronese, S.; Wang, T. L.; Weingart, J. D.; Wolfgang, C. L.; Wood, L. D.; Xing, D.; Hruban, R. H.; Wu, J.; Allen, P. J.; Schmidt, C. M.; Choti, M. A.; Velculescu, V. E.; Kinzler, K. W.; Vogelstein, B.; Papadopoulos, N.; Diaz, L. A. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci. Transl. Med.* **2014**, *6* (224), 224ra24. <https://doi.org/10.1126/SCITRANSLMED.3007094>.
 - (28) Lone, S. N.; Nisar, S.; Masoodi, T.; Singh, M.; Rizwan, A.; Hashem, S.; El-Rifai, W.; Bedognetti, D.; Batra, S. K.; Haris, M.; Bhat, A. A.; Macha, M. A. Liquid Biopsy: A Step Closer to Transform Diagnosis, Prognosis and Future of Cancer Treatments. *Mol. Cancer* **2022**, *21* (1), 1–22. <https://doi.org/10.1186/S12943-022-01543-7>.
 - (29) Li, X.; Ye, M.; Zhang, W.; Tan, D.; Jaffrezic-Renault, N.; Yang, X.; Guo, Z. Liquid Biopsy of Circulating Tumor DNA and Biosensor Applications. *Biosens. Bioelectron.* **2019**, *126*, 596–607. <https://doi.org/10.1016/J.BIOS.2018.11.037>.
 - (30) Underhill, H. R. Leveraging the Fragment Length of Circulating Tumour DNA to Improve Molecular Profiling of Solid Tumour Malignancies with Next-Generation Sequencing: A Pathway to Advanced Non-Invasive Diagnostics in Precision Oncology? *Mol. Diagnosis Ther.* **2021**, *25* (4), 389–408. <https://doi.org/10.1007/S40291-021-00534-6>.
 - (31) Fleischhacker, M.; Schmidt, B. Circulating Nucleic Acids (CNAs) and Cancer—A Survey. *Biochim. Biophys. Acta - Rev. Cancer* **2007**, *1775* (1), 181–232. <https://doi.org/10.1016/J.BBCAN.2006.10.001>.
 - (32) Sánchez-Herrero, E.; Serna-Blasco, R.; Robado de Lope, L.; González-Rumayor, V.; Romero, A.; Provencio, M. Circulating Tumor DNA as a Cancer Biomarker: An Overview of Biological Features and Factors That May Impact on CtDNA Analysis. *Front. Oncol.* **2022**, *12*, 3410. <https://doi.org/10.3389/FONC.2022.943253/BIBTEX>.

- (33) Kong, Y. W.; Ferland-McCollough, D.; Jackson, T. J.; Bushell, M. MicroRNAs in Cancer Management. *Lancet Oncol.* **2012**, *13* (6), e249–e258. [https://doi.org/10.1016/S1470-2045\(12\)70073-6](https://doi.org/10.1016/S1470-2045(12)70073-6).
- (34) Peng, Y.; Croce, C. M. The Role of MicroRNAs in Human Cancer. *Signal Transduct. Target. Ther.* **2016**, *1* (1), 1–9. <https://doi.org/10.1038/sigtrans.2015.4>.
- (35) Si, M. L.; Zhu, S.; Wu, H.; Lu, Z.; Wu, F.; Mo, Y. Y. MiR-21-Mediated Tumor Growth. *Oncogene* **2006**, *26* (19), 2799–2803. <https://doi.org/10.1038/sj.onc.1210083>.
- (36) Iorio, M. V.; Ferracin, M.; Liu, C. G.; Veronese, A.; Spizzo, R.; Sabbioni, S.; Magri, E.; Pedriali, M.; Fabbri, M.; Campiglio, M.; Ménard, S.; Palazzo, J. P.; Rosenberg, A.; Musiani, P.; Volinia, S.; Nenci, I.; Calin, G. A.; Querzoli, P.; Negrini, M.; Croce, C. M. MicroRNA Gene Expression Deregulation in Human Breast Cancer. *Cancer Res.* **2005**, *65* (16), 7065–7070. <https://doi.org/10.1158/0008-5472.CAN-05-1783>.
- (37) Schetter, A. J.; Leung, S. Y.; Sohn, J. J.; Zanetti, K. A.; Bowman, E. D.; Yanaihara, N.; Yuen, S. T.; Chan, T. L.; Kwong, D. L. W.; Au, G. K. H.; Liu, C. G.; Calin, G. A.; Croce, C. M.; Harris, C. C. MicroRNA Expression Profiles Associated with Prognosis and Therapeutic Outcome in Colon Adenocarcinoma. *JAMA* **2008**, *299* (4), 425–436. <https://doi.org/10.1001/JAMA.299.4.425>.
- (38) Fridrichova, I.; Zmetakova, I. MicroRNAs Contribute to Breast Cancer Invasiveness. *Cells* **2019**, *8* (11). <https://doi.org/10.3390/CELLS8111361>.
- (39) Rupaimoole, R.; Slack, F. J. MicroRNA Therapeutics: Towards a New Era for the Management of Cancer and Other Diseases. *Nat. Rev. Drug Discov.* **2017**, *16* (3), 203–222. <https://doi.org/10.1038/nrd.2016.246>.
- (40) Hanna, J.; Hossain, G. S.; Kocerha, J. The Potential for MicroRNA Therapeutics and Clinical Research. *Front. Genet.* **2019**, *10* (MAY), 478. <https://doi.org/10.3389/FGENE.2019.00478/BIBTEX>.
- (41) Kasinski, A. L.; Kelnar, K.; Stahlhut, C.; Orellana, E.; Zhao, J.; Shimer, E.; Dysart, S.; Chen, X.; Bader, A. G.; Slack, F. J. A Combinatorial MicroRNA Therapeutics Approach to Suppressing Non-Small Cell Lung Cancer. *Oncogene* **2015**, *34* (27), 3547. <https://doi.org/10.1038/ONC.2014.282>.
- (42) Wu, J.; Huang, J.; Kuang, S.; Chen, J.; Li, X.; Chen, B.; Wang, J.; Cheng, D.; Shuai, X. Synergistic MicroRNA Therapy in Liver Fibrotic Rat Using MRI-Visible Nanocarrier Targeting Hepatic Stellate Cells. *Adv. Sci. (Weinheim, Baden-Wurttemberg, Ger.)* **2019**, *6* (5). <https://doi.org/10.1002/ADVS.201801809>.
- (43) X, S.; Y, H.; Y, L.; JL, C.; XQ, W.; X, M.; HF, X. The Polymorphism of Rs6505162 in the MIR423 Coding Region and Recurrent Pregnancy Loss. *Reproduction* **2015**, *150* (1), 65–76. <https://doi.org/10.1530/REP-15-0007>.
- (44) Ryan, B. M.; Robles, A. I.; Harris, C. C. Genetic Variation in MicroRNA Networks: The Implications for Cancer Research. *Nat. Rev. Cancer* **2010**, *10* (6), 389–402. <https://doi.org/10.1038/NRC2867>.

-
- (45) Pawlina-Tyszko, K.; Semik-Gurgul, E.; Gurgul, A.; Oczkowicz, M.; Szmatoła, T.; Bugno-Poniewierska, M. Application of the Targeted Sequencing Approach Reveals the Single Nucleotide Polymorphism (SNP) Repertoire in MicroRNA Genes in the Pig Genome. *Sci. Rep.* **2021**, *11* (1). <https://doi.org/10.1038/S41598-021-89363-5>.
- (46) Ouyang, T.; Liu, Z.; Han, Z.; Ge, Q. MicroRNA Detection Specificity: Recent Advances and Future Perspective. *Anal. Chem.* **2019**, *91* (5), 3179–3186. <https://doi.org/10.1021/ACS.ANALCHEM.8B05909>.
- (47) Huerta, M.; Roselló, S.; Sabater, L.; Ferrer, A.; Tarazona, N.; Roda, D.; Gambardella, V.; Alfaro-Cervelló, C.; Garcés-Albir, M.; Cervantes, A.; Ibarrola-Villava, M. Circulating Tumor DNA Detection by Digital-Droplet PCR in Pancreatic Ductal Adenocarcinoma: A Systematic Review. *Cancers (Basel)*. **2021**, *13* (5), 1–14. <https://doi.org/10.3390/CANCERS13050994>.
- (48) Chen, C.; Ridzon, D. A.; Broomer, A. J.; Zhou, Z.; Lee, D. H.; Nguyen, J. T.; Barbisin, M.; Xu, N. L.; Mahuvakar, V. R.; Andersen, M. R.; Lao, K. Q.; Livak, K. J.; Guegler, K. J. Real-Time Quantification of MicroRNAs by Stem-Loop RT-PCR. *Nucleic Acids Res.* **2005**, *33* (20). <https://doi.org/10.1093/NAR/GNI178>.
- (49) Chen, C.; Tan, R.; Wong, L.; Fekete, R.; Halsey, J. Quantitation of MicroRNAs by Real-Time RT-QPCR. *Methods Mol. Biol.* **2011**, *687*, 113–134. https://doi.org/10.1007/978-1-60761-944-4_8.
- (50) Yi, M.; Liao, Z.; Deng, L.; Xu, L.; Tan, Y.; Liu, K.; Chen, Z.; Zhang, Y. High Diagnostic Value of MiRNAs for NSCLC: Quantitative Analysis for Both Single and Combined MiRNAs in Lung Cancer. *Ann. Med.* **2021**, *53* (1), 2178. <https://doi.org/10.1080/07853890.2021.2000634>.
- (51) Saelee, S. L.; Lovejoy, A. F.; Hinzmann, B.; Mayol, K.; Huynh, S.; Harrell, A.; Lefkowitz, J.; Deodhar, N.; Garcia-Montoya, G.; Yaung, S. J.; Klass, D. M. Quantitative PCR-Based Method to Assess Cell-Free DNA Quality, Adjust Input Mass, and Improve Next-Generation Sequencing Assay Performance. *J. Mol. Diagnostics* **2022**, *24* (6), 566–575. <https://doi.org/10.1016/J.JMOLDX.2022.02.005>.
- (52) Vogelstein, B.; Kinzler, K. W. Digital PCR. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (16), 9236–9241. <https://doi.org/10.1073/PNAS.96.16.9236/ASSET/6FC5DE41-B0CD-47D1-B568-542383577EE3/ASSETS/GRAPHIC/PQ1692350005.JPEG>.
- (53) Minato, T.; Ito, S.; Li, B.; Fujimori, H.; Mochizuki, M.; Yamaguchi, K.; Tamai, K.; Shimada, M.; Tokunaga, H.; Shigeta, S.; Sato, I.; Shima, H.; Yamada, H.; Yaegashi, N.; Yasuda, J. Liquid Biopsy with Droplet Digital PCR Targeted to Specific Mutations in Plasma Cell-Free Tumor DNA Can Detect Ovarian Cancer Recurrence Earlier than CA125. *Gynecol. Oncol. Reports* **2021**, *38*, 100847. <https://doi.org/10.1016/J.GORE.2021.100847>.
- (54) Thress, K. S.; Brant, R.; Carr, T. H.; Dearden, S.; Jenkins, S.; Brown, H.; Hammett, T.; Cantarini, M.; Barrett, J. C. EGFR Mutation Detection in CtDNA from NSCLC Patient Plasma: A Cross-Platform Comparison of Leading Technologies to Support the Clinical Development of AZD9291. *Lung Cancer* **2015**, *90* (3), 509–515.
-

- <https://doi.org/10.1016/J.LUNGCAN.2015.10.004>.
- (55) Salk, J. J.; Schmitt, M. W.; Loeb, L. A. Enhancing the Accuracy of Next-Generation Sequencing for Detecting Rare and Subclonal Mutations. *Nat. Rev. Genet.* 2018 195 **2018**, 19 (5), 269–285. <https://doi.org/10.1038/nrg.2017.117>.
- (56) Illumina. *Indexed Sequencing Overview Guide*. https://jp.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/indexed-sequencing-overview-guide-15057455-08.pdf (accessed 2023-01-25).
- (57) Yoshinami, T.; Kagara, N.; Motooka, D.; Nakamura, S.; Miyake, T.; Tanei, T.; Naoi, Y.; Shimoda, M.; Shimazu, K.; Kim, S. J.; Noguchi, S. Detection of CtDNA with Personalized Molecular Barcode NGS and Its Clinical Significance in Patients with Early Breast Cancer. *Transl. Oncol.* **2020**, 13 (8), 100787. <https://doi.org/10.1016/J.TRANON.2020.100787>.
- (58) Khamina, K.; Diendorfer, A. B.; Skalicky, S.; Weigl, M.; Pultar, M.; Krammer, T. L.; Fournier, C. A.; Schofield, A. L.; Otto, C.; Smith, A. T.; Buchtele, N.; Schoergenhofer, C.; Jilma, B.; Frank, B. J. H.; Hofstaetter, J. G.; Grillari, R.; Grillari, J.; Ruprecht, K.; Goldring, C. E.; Rehrauer, H.; Glaab, W. E.; Hackl, M. A MicroRNA Next-Generation-Sequencing Discovery Assay (MiND) for Genome-Scale Analysis and Absolute Quantitation of Circulating MicroRNA Biomarkers. *Int. J. Mol. Sci.* **2022**, 23 (3). <https://doi.org/10.3390/IJMS23031226/S1>.
- (59) Newman, A. M.; Bratman, S. V.; To, J.; Wynne, J. F.; Eclov, N. C. W.; Modlin, L. A.; Liu, C. L.; Neal, J. W.; Wakelee, H. A.; Merritt, R. E.; Shrager, J. B.; Loo, B. W.; Alizadeh, A. A.; Diehn, M. An Ultrasensitive Method for Quantitating Circulating Tumor DNA with Broad Patient Coverage. *Nat. Med.* **2014**, 20 (5), 548–554. <https://doi.org/10.1038/NM.3519>.
- (60) Kinde, I.; Wu, J.; Papadopoulos, N.; Kinzler, K. W.; Vogelstein, B. Detection and Quantification of Rare Mutations with Massively Parallel Sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, 108 (23), 9530–9535. https://doi.org/10.1073/PNAS.1105422108/SUPPL_FILE/PNAS.201105422SI.PDF.
- (61) Kato, R.; Hayashi, H.; Sakai, K.; Suzuki, S.; Haratani, K.; Takahama, T.; Tanizaki, J.; Nonagase, Y.; Tanaka, K.; Yoshida, T.; Takeda, M.; Yonesaka, K.; Kaneda, H.; Nishio, K.; Nakagawa, K. CAPP-Seq Analysis of Circulating Tumor DNA from Patients with EGFR T790M-Positive Lung Cancer after Osimertinib. *Int. J. Clin. Oncol.* **2021**, 26 (9), 1628–1639. <https://doi.org/10.1007/S10147-021-01947-3>.
- (62) Fredebohm, J.; Mehnert, D. H.; Löber, A. K.; Holtrup, F.; van Rahden, V.; Angenendt, P.; Diehl, F. Detection and Quantification of KIT Mutations in CtDNA by Plasma Safe-SeqS. *Adv. Exp. Med. Biol.* **2016**, 924, 187–189. https://doi.org/10.1007/978-3-319-42044-8_34.
- (63) Tie, J.; Kinde, I.; Wang, Y.; Wong, H. L.; Roebert, J.; Christie, M.; Tacey, M.; Wong, R.; Singh, M.; Karapetis, C. S.; Desai, J.; Tran, B.; Strausberg, R. L.; Diaz, L. A.; Papadopoulos, N.; Kinzler, K. W.; Vogelstein, B.; Gibbs, P. Circulating Tumor DNA as an Early Marker of Therapeutic Response in Patients with Metastatic Colorectal Cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2015**, 26 (8), 1715–1722. <https://doi.org/10.1093/ANNONC/MDV177>.

-
- (64) Zhang, Y.; Sui, J.; Shen, X.; Li, C.; Yao, W.; Hong, W.; Peng, H.; Pu, Y.; Yin, L.; Liang, G. Differential Expression Profiles of MicroRNAs as Potential Biomarkers for the Early Diagnosis of Lung Cancer. *Oncol. Rep.* **2017**, *37* (6), 3543–3553. <https://doi.org/10.3892/OR.2017.5612/HTML>.
- (65) Liang, F.; Yang, M.; Tong, N.; Fang, J.; Pan, Y.; Li, J.; Zhang, X. Identification of Six Key MiRNAs Associated with Breast Cancer through Screening Large-Scale Microarray Data. *Oncol. Lett.* **2018**, *16* (4), 4159. <https://doi.org/10.3892/OL.2018.9175>.
- (66) Qiu, X.; Wang, P.; Cao, Z. Hybridization Chain Reaction Modulated DNA-Hosted Silver Nanoclusters for Fluorescent Identification of Single Nucleotide Polymorphisms in the Let-7 MiRNA Family. *Biosens. Bioelectron.* **2014**, *60*, 351–357. <https://doi.org/10.1016/J.BIOS.2014.04.040>.
- (67) Caputo, T. M.; Battista, E.; Netti, P. A.; Causa, F. Supramolecular Microgels with Molecular Beacons at the Interface for Ultrasensitive, Amplification-Free, and SNP-Selective MiRNA Fluorescence Detection. *ACS Appl. Mater. Interfaces* **2019**, *11* (19), 17147–17156. <https://doi.org/10.1021/ACSAMI.8B22635>.
- (68) Roy, S.; Soh, J. H.; Ying, J. Y. A Microarray Platform for Detecting Disease-Specific Circulating MiRNA in Human Serum. *Biosens. Bioelectron.* **2016**, *75*, 238–246. <https://doi.org/10.1016/J.BIOS.2015.08.039>.
- (69) Jia, H. Y.; Zhao, H. L.; Wang, T.; Chen, P. R.; Yin, B. C.; Ye, B. C. A Programmable and Sensitive CRISPR/Cas12a-Based MicroRNA Detection Platform Combined with Hybridization Chain Reaction. *Biosens. Bioelectron.* **2022**, *211*. <https://doi.org/10.1016/J.BIOS.2022.114382>.
- (70) Nimse, S. B.; Song, K.; Sonawane, M. D.; Sayyed, D. R.; Kim, T. Immobilization Techniques for Microarray: Challenges and Applications. *Sensors (Basel)*. **2014**, *14* (12), 22208. <https://doi.org/10.3390/S141222208>.
- (71) Dufva, M.; Petersen, J.; Stoltenborg, M.; Birgens, H.; Christensen, C. B. V. Detection of Mutations Using Microarrays of Poly(C)10-Poly(T)10 Modified DNA Probes Immobilized on Agarose Films. *Anal. Biochem.* **2006**, *352* (2), 188–197. <https://doi.org/10.1016/J.AB.2006.03.008>.
- (72) Sabourin, D.; Petersen, J.; Snakenborg, D.; Brivio, M.; Gudnadson, H.; Wolff, A.; Dufva, M. Microfluidic DNA Microarrays in PMMA Chips: Streamlined Fabrication via Simultaneous DNA Immobilization and Bonding Activation by Brief UV Exposure. *Biomed. Microdevices* **2010**, *12* (4), 673–681. <https://doi.org/10.1007/S10544-010-9420-7>.
- (73) Kim, Y. J.; Hosokawa, K.; Maeda, M. Sensitivity Enhancement of MicroRNA Detection Using a Power-Free Microfluidic Chip. *Anal. Sci.* **2019**, *35* (11), 1227–1236. <https://doi.org/10.2116/ANALSCI.19P211>.
- (74) Fixe, F.; Dufva, M.; Telleman, P.; Christensen, C. B. Functionalization of Poly(Methyl Methacrylate) (PMMA) as a Substrate for DNA Microarrays. *Nucleic Acids Res.* **2004**, *32* (1).
-

- <https://doi.org/10.1093/NAR/GNG157>.
- (75) Gudnason, H.; Dufva, M.; Bang, D. D.; Wolff, A. An Inexpensive and Simple Method for Thermally Stable Immobilization of DNA on an Unmodified Glass Surface: UV Linking of Poly(T) 10-Poly(C) 10-Tagged DNA Probes. *Biotechniques* **2008**, 45 (3), 261–271. <https://doi.org/10.2144/000112905/ASSET/IMAGES/LARGE/FIGURE3.JPEG>.
- (76) Domljanovic, I.; Taskova, M.; Miranda, P.; Weber, G.; Astakhova, K. Optical and Theoretical Study of Strand Recognition by Nucleic Acid Probes. *Commun. Chem.* **2020** 31 **2020**, 3 (1), 1–13. <https://doi.org/10.1038/s42004-020-00362-5>.
- (77) Haslam, N. J.; Whiteford, N. E.; Weber, G.; Prügel-Bennett, A.; Essex, J. W.; Neylon, C. Optimal Probe Length Varies for Targets with High Sequence Variation: Implications for Probe Library Design for Resequencing Highly Variable Genes. *PLoS One* **2008**, 3 (6), e2500. <https://doi.org/10.1371/JOURNAL.PONE.0002500>.
- (78) Muniz, M. I.; Bustos, A. H.; Slott, S.; Astakhova, K.; Weber, G. Cation Valence Dependence of Hydrogen Bond and Stacking Potentials in DNA Mesoscopic Models. *Biophys. Chem.* **2022**, 106949. <https://doi.org/10.1016/J.BPC.2022.106949>.
- (79) Tan, Z. J.; Chen, S. J. Nucleic Acid Helix Stability: Effects of Salt Concentration, Cation Valence and Size, and Chain Length. *Biophys. J.* **2006**, 90 (4), 1175. <https://doi.org/10.1529/BIOPHYSJ.105.070904>.
- (80) Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Olsen, C. E.; Wengel, J. LNA (Locked Nucleic Acids): Synthesis of Adenin, Cytosine, Guanine, 5-Methylcytosine, Thymine and Uracil Bicyclonucleoside Monomers, Oligomerisation, and Unprecedented Nucleic Acid Recognition. *Tetrahedron* **1998**, 54 (14), 3607–3630. [https://doi.org/10.1016/S0040-4020\(98\)00094-5](https://doi.org/10.1016/S0040-4020(98)00094-5).
- (81) Obika, S.; Nanbu, D.; Hari, Y.; Morio, K. I.; In, Y.; Ishida, T.; Imanishi, T. Synthesis of 2'-O,4'-C-Methyleneuridine and -Cytidine. Novel Bicyclic Nucleosides Having a Fixed C3, -Endo Sugar Puckering. *Tetrahedron Lett.* **1997**, 38 (50), 8735–8738. [https://doi.org/10.1016/S0040-4039\(97\)10322-7](https://doi.org/10.1016/S0040-4039(97)10322-7).
- (82) Wengel, J.; Petersen, M.; Nielsen, K. E.; Jensen, G. A.; Håkansson, A. E.; Kumar, R.; Sørensen, M. D.; Rajwanshi, V. K.; Bryld, T.; Jacobsen, J. P. LNA (Locked Nucleic Acid) and the Diastereoisomeric Alpha-L-LNA: Conformational Tuning and High-Affinity Recognition of DNA/RNA Targets. *Nucleosides. Nucleotides Nucleic Acids* **2001**, 20 (4–7), 389–396. <https://doi.org/10.1081/NCN-100002312>.
- (83) Tolstrup, N.; Nielsen, P. S.; Kolberg, J. G.; Frankel, A. M.; Vissing, H.; Kauppinen, S. OligoDesign: Optimal Design of LNA (Locked Nucleic Acid) Oligonucleotide Capture Probes for Gene Expression Profiling. *Nucleic Acids Res.* **2003**, 31 (13), 3758. <https://doi.org/10.1093/NAR/GKG580>.
- (84) Renneberg, D.; Leumann, C. J. Watson-Crick Base-Pairing Properties of Tricyclo-DNA. **2002**. <https://doi.org/10.1021/JA025569>.
- (85) Seth, P. P.; Siwkowski, A.; Allerson, C. R.; Vasquez, G.; Lee, S.; Prakash, T. P.; Kinberger,

-
- G.; Migawa, M. T.; Gaus, H.; Bhat, B.; Swayze, E. E. Design, Synthesis and Evaluation of Constrained Methoxyethyl (CMOE) and Constrained Ethyl (CEt) Nucleoside Analogs. *Nucleic Acids Symp. Ser. (Oxf)*. **2008**, No. 52, 553–554. <https://doi.org/10.1093/NASS/NRN280>.
- (86) Morita, K.; Hasegawa, C.; Kaneko, M.; Tsutsumi, S.; Sone, J.; Ishikawa, T.; Imanishi, T.; Koizumi, M. 2'-O,4'-C-Ethylene-Bridged Nucleic Acids (ENA) with Nuclease-Resistance and High Affinity for RNA. *Nucleic Acids Res. Suppl.* **2001**, No. 1, 241–242. <https://doi.org/10.1093/NASS/1.1.241>.
- (87) Agrawal, S.; Gait, M. J. History and Development of Nucleotide Analogues in Nucleic Acids Drugs. *RSC Drug Discov. Ser.* **2019**, 2019-January (68), 1–21. <https://doi.org/10.1039/9781788015714-00001>.
- (88) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. Peptide Nucleic Acids (PNA). Oligonucleotide Analogues with an Achiral Peptide Backbone. *J. Am. Chem. Soc.* **1992**, 114 (5), 1895–1897. https://doi.org/10.1021/JA00031A062/SUPPL_FILE/JA00031A062_SI_001.PDF.
- (89) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide. *Science* **1991**, 254 (5037), 1497–1500. <https://doi.org/10.1126/SCIENCE.1962210>.
- (90) Cadoni, E.; Manicardi, A.; Madder, A. PNA-Based MicroRNA Detection Methodologies. *Molecules* **2020**, 25 (6). <https://doi.org/10.3390/MOLECULES25061296>.
- (91) Tian, K.; He, Z.; Wang, Y.; Chen, S. J.; Gu, L. Q. Designing a Polycationic Probe for Simultaneous Enrichment and Detection of MicroRNAs in a Nanopore. *ACS Nano* **2013**, 7 (5), 3962–3969. https://doi.org/10.1021/NN305789Z/SUPPL_FILE/NN305789Z_SI_001.PDF.
- (92) Bruylants, G.; Bocconcelli, M.; Snoussi, K.; Bartik, K. Comparison of the Thermodynamics and Base-Pair Dynamics of a Full LNA:DNA Duplex and of the Isosequential DNA:DNA Duplex. *Biochemistry* **2009**, 48 (35), 8473–8482. <https://doi.org/10.1021/BI900615Z>.
- (93) Owczarzy, R.; You, Y.; Groth, C. L.; Tataurov, A. V. Stability and Mismatch Discrimination of Locked Nucleic Acid-DNA Duplexes. *Biochemistry* **2011**, 50 (43), 9352–9367. <https://doi.org/10.1021/BI200904E>.
- (94) Hughesman, C. B.; Turner, R. F. B.; Haynes, C. A. Role of the Heat Capacity Change in Understanding and Modeling Melting Thermodynamics of Complementary Duplexes Containing Standard and Nucleobase-Modified LNA. *Biochemistry* **2011**, 50 (23), 5354–5368. <https://doi.org/10.1021/BI200223S>.
- (95) Fakhfakh, K.; Marais, O.; Cheng, X. B. J.; Castañeda, J. R.; Hughesman, C. B.; Haynes, C. Molecular Thermodynamics of LNA:LNA Base Pairs and the Hyperstabilizing Effect of 5'-Proximal LNA:DNA Base Pairs. *AIChE J.* **2015**, 61 (9), 2711–2731. <https://doi.org/10.1002/AIC.14916>.
- (96) McTigue, P. M.; Peterson, R. J.; Kahn, J. D. Sequence-Dependent Thermodynamic
-

- Parameters for Locked Nucleic Acid (LNA)-DNA Duplex Formation. *Biochemistry* **2004**, 43 (18), 5388–5405. <https://doi.org/10.1021/BI035976D>.
- (97) You, Y.; Moreira, B. G.; Behlke, M. A.; Owczarzy, R. Design of LNA Probes That Improve Mismatch Discrimination. *Nucleic Acids Res.* **2006**, 34 (8). <https://doi.org/10.1093/NAR/GKL175>.
- (98) Huguet, J. M.; Ribezzi-Crivellari, M.; Bizarro, C. V.; Ritort, F. Derivation of Nearest-Neighbor DNA Parameters in Magnesium from Single Molecule Experiments. *Nucleic Acids Res.* **2017**, 45 (22), 12921. <https://doi.org/10.1093/NAR/GKX1161>.
- (99) Ferreira, I.; Slott, S.; Astakhova, K.; Weber, G. Complete Mesoscopic Parameterization of Single LNA Modifications in DNA Applied to Oncogene Probe Design. *J. Chem. Inf. Model.* **2021**, 61 (7), 3615–3624. <https://doi.org/10.1021/ACS.JCIM.1C00470>.
- (100) Merrifield, B. Solid Phase Synthesis. *Science* (80-.). **1986**, 232 (4748), 341–347. <https://doi.org/10.1126/SCIENCE.3961484/ASSET/BF560F8E-6CC5-4F65-9711-8E608B82A0A7/ASSETS/SCIENCE.3961484.FP.PNG>.
- (101) Beaucage, S. L.; Caruthers, M. H. Deoxynucleoside Phosphoramidites—A New Class of Key Intermediates for Deoxypolynucleotide Synthesis. *Tetrahedron Lett.* **1981**, 22 (20), 1859–1862. [https://doi.org/10.1016/S0040-4039\(01\)90461-7](https://doi.org/10.1016/S0040-4039(01)90461-7).
- (102) McBride, L. J.; Caruthers, M. H. An Investigation of Several Deoxynucleoside Phosphoramidites Useful for Synthesizing Deoxyoligonucleotides. *Tetrahedron Lett.* **1983**, 24 (3), 245–248. [https://doi.org/10.1016/S0040-4039\(00\)81376-3](https://doi.org/10.1016/S0040-4039(00)81376-3).
- (103) Carr, P. A.; Park, J. S.; Lee, Y. J.; Yu, T.; Zhang, S.; Jacobson, J. M. Protein-Mediated Error Correction for de Novo DNA Synthesis. *Nucleic Acids Res.* **2004**, 32 (20). <https://doi.org/10.1093/NAR/GNH160>.
- (104) Pon, R. T.; Yu, S. Tandem Oligonucleotide Synthesis Using Linker Phosphoramidites. *Nucleic Acids Res.* **2005**, 33 (6), 1940–1948. <https://doi.org/10.1093/NAR/GKI333>.
- (105) Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. Light-Directed, Spatially Addressable Parallel Chemical Synthesis. *Science* **1991**, 251 (4995), 767–773. <https://doi.org/10.1126/SCIENCE.1990438>.
- (106) Pease, A. C.; Solas, D.; Sullivan, E. J.; Cronin, M. T.; Holmes, C. P.; Fodor, S. P. A. Light-Generated Oligonucleotide Arrays for Rapid DNA Sequence Analysis. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, 91 (11), 5022–5026. <https://doi.org/10.1073/PNAS.91.11.5022>.
- (107) Singh-Gasson, S.; Green, R. D.; Yue, Y.; Nelson, C.; Blattner, F.; Sussman, M. R.; Cerrina, F. Maskless Fabrication of Light-Directed Oligonucleotide Microarrays Using a Digital Micromirror Array. *Nat. Biotechnol.* **1999**, 17 (10), 974–978. <https://doi.org/10.1038/13664>.
- (108) Gao, X.; LeProust, E.; Zhang, H.; Srivannavit, O.; Gulari, E.; Yu, P.; Nishiguchi, C.; Xiang, Q.; Zhou, X. A Flexible Light-Directed DNA Chip Synthesis Gated by Deprotection Using Solution Photogenerated Acids. *Nucleic Acids Res.* **2001**, 29 (22), 4744. <https://doi.org/10.1093/NAR/29.22.4744>.

-
- (109) LeProust, E. M.; Peck, B. J.; Spirin, K.; McCuen, H. B.; Moore, B.; Namsaraev, E.; Caruthers, M. H. Synthesis of High-Quality Libraries of Long (150mer) Oligonucleotides by a Novel Depurination Controlled Process. *Nucleic Acids Res.* **2010**, *38* (8), 2522–2540. <https://doi.org/10.1093/NAR/GKQ163>.
- (110) Kuhn, P.; Wagner, K.; Heil, K.; Liss, M.; Netuschil, N. Next Generation Gene Synthesis: From Microarrays to Genomes. *Eng. Life Sci.* **2017**, *17* (1), 6. <https://doi.org/10.1002/ELSC.201600121>.
- (111) Uhd, J.; Miotke, L.; Ji, H. P.; Dunaeva, M.; Pruijn, G. J. M.; Jørgensen, C. D.; Kristoffersen, E. L.; Birkedal, V.; Yde, C. W.; Nielsen, F. C.; Hansen, J.; Astakhova, K. Ultra-Fast Detection and Quantification of Nucleic Acids by Amplification-Free Fluorescence Assay. *Analyst* **2020**, *145* (17), 5836–5844. <https://doi.org/10.1039/D0AN00676A>.
- (112) Fajardo, P.; Taskova, M.; Martín-Serrano, M. A.; Hansen, J.; Slott, S.; Jakobsen, A. K.; Wibom, M. L.; Salegi, B.; Muñoz, A.; Barbachano, A.; Sharma, A.; Gubatan, J. M.; Habtezion, A.; Sanz-Ezquerro, J. J.; Astakhova, K.; Cuenda, A. P38 γ and P38 δ as Biomarkers in the Interplay of Colon Cancer and Inflammatory Bowel Diseases. *Cancer Commun. (London, England)* **2022**, *42* (9), 897–901. <https://doi.org/10.1002/CAC2.12331>.
- (113) Setten, R. L.; Dowdy, S. F. DNA/RNA Heteroduplex Oligonucleotides: An Unanticipated Twist in the Delivery of ASOs. **2022**. <https://doi.org/10.1016/j.omtn.2022.06.013>.
- (114) Chen, M.; Zhao, H. Next-Generation Sequencing in Liquid Biopsy: Cancer Screening and Early Detection. *Hum. Genomics* **2019**, *13* (1), 34. <https://doi.org/10.1186/S40246-019-0220-8/TABLES/1>.
- (115) Mamanova, L.; Coffey, A. J.; Scott, C. E.; Kozarewa, I.; Turner, E. H.; Kumar, A.; Howard, E.; Shendure, J.; Turner, D. J. Target-Enrichment Strategies for next-Generation Sequencing. *Nat. Methods* **2010**, *7* (2), 111–118. <https://doi.org/10.1038/nmeth.1419>.
- (116) Aird, D.; Ross, M. G.; Chen, W. S.; Danielsson, M.; Fennell, T.; Russ, C.; Jaffe, D. B.; Nusbaum, C.; Gnirke, A. Analyzing and Minimizing PCR Amplification Bias in Illumina Sequencing Libraries. *Genome Biol.* **2011**, *12* (2), 1–14. <https://doi.org/10.1186/GB-2011-12-2-R18/FIGURES/7>.
- (117) Hong, B. J.; Oh, S. J.; Youn, T. O.; Kwon, S. H.; Park, J. W. Nanoscale-Controlled Spacing Provides DNA Microarrays with the SNP Discrimination Efficiency in Solution Phase. *Langmuir* **2005**, *21* (10), 4257–4261. <https://doi.org/10.1021/LA046951Y>.
- (118) Oh, S. J.; Ju, J.; Kim, B. C.; Ko, E.; Hong, B. J.; Park, J. G.; Park, J. W.; Choi, K. Y. DNA Microarrays on a Dendron-Modified Surface Improve Significantly the Detection of Single Nucleotide Variations in the P53 Gene. *Nucleic Acids Res.* **2005**, *33* (10), 1–8. <https://doi.org/10.1093/NAR/GNI087>.
- (119) Bagheri, P.; Sharifi, M.; Ghadiri, A. Downregulation of MIR100HG Induces Apoptosis in Human Megakaryoblastic Leukemia Cells. *Indian J. Hematol. Blood Transfus.* **2021**, *37* (2), 232–239. <https://doi.org/10.1007/S12288-020-01324-6>.
-

- (120) Zhang, K.; Zheludev, I. N.; Hagey, R. J.; Haslecker, R.; Hou, Y. J.; Kretsch, R.; Pintilie, G. D.; Rangan, R.; Kladwang, W.; Li, S.; Wu, M. T. P.; Pham, E. A.; Bernardin-Souibgui, C.; Baric, R. S.; Sheahan, T. P.; D'Souza, V.; Glenn, J. S.; Chiu, W.; Das, R. Cryo-EM and Antisense Targeting of the 28-KDa Frameshift Stimulation Element from the SARS-CoV-2 RNA Genome. *Nat. Struct. Mol. Biol.* **2021**, *28* (9), 747–754. <https://doi.org/10.1038/s41594-021-00653-y>.
- (121) Lim, K. R. Q.; Maruyama, R.; Echigoya, Y.; Nguyen, Q.; Zhang, A.; Khawaja, H.; Chandra, S. Sen; Jones, T.; Jones, P.; Chen, Y. W.; Yokota, T. Inhibition of DUX4 Expression with Antisense LNA Gapmers as a Therapy for Facioscapulohumeral Muscular Dystrophy. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (28), 16509–16515. <https://doi.org/10.1073/PNAS.1909649117>.
- (122) Josefsen, M. H.; Löfström, C.; Sommer, H. M.; Hoorfar, J. Diagnostic PCR: Comparative Sensitivity of Four Probe Chemistries. *Mol. Cell. Probes* **2009**, *23* (3–4), 201–203. <https://doi.org/10.1016/J.MCP.2009.02.003>.
- (123) Astakhova, I. V.; Ustinov, A. V.; Korshun, V. A.; Wengel, J. LNA for Optimization of Fluorescent Oligonucleotide Probes: Improved Spectral Properties and Target Binding. *Bioconjug. Chem.* **2011**, *22* (4), 533–539. <https://doi.org/10.1021/BC1005027>.
- (124) Weber, G.; Essex, J. W.; Neylon, C. Probing the Microscopic Flexibility of DNA from Melting Temperatures. *Nat. Phys.* **2009**, *5* (10), 769–773.
- (125) Oliveira, L. M.; Long, A. S.; Brown, T.; Fox, K. R.; Weber, G. Melting Temperature Measurement and Mesoscopic Evaluation of Single, Double and Triple DNA Mismatches. *Chem. Sci.* **2020**, *11* (31), 8273. <https://doi.org/10.1039/D0SC01700K>.
- (126) Weber, G.; Haslam, N.; Essex, J. W.; Neylon, C. Thermal Equivalence of DNA Duplexes for Probe Design. *J. Phys. Condens. Matter* **2008**, *21* (3), 034106. <https://doi.org/10.1088/0953-8984/21/3/034106>.
- (127) de Oliveira Martins, E.; Basílio Barbosa, V.; Weber, G. DNA/RNA Hybrid Mesoscopic Model Shows Strong Stability Dependence with Deoxypyrimidine Content and Stacking Interactions Similar to RNA/RNA. *Chem. Phys. Lett.* **2019**, *715*, 14–19. <https://doi.org/10.1016/J.CPLETT.2018.11.015>.
- (128) Sugimoto, N.; Nakano, S. ichi; Katoh, M.; Matsumura, A.; Nakamuta, H.; Ohmichi, T.; Yoneyama, M.; Sasaki, M. Thermodynamic Parameters to Predict Stability of RNA/DNA Hybrid Duplexes. *Biochemistry* **1995**, *34* (35), 11211–11216. <https://doi.org/10.1021/BI00035A029>.
- (129) Amarante, T. D.; Weber, G. Evaluating Hydrogen Bonds and Base Stacking of Single, Tandem and Terminal GU Mismatches in RNA with a Mesoscopic Model. *J. Chem. Inf. Model.* **2016**, *56* (1), 101–109. <https://doi.org/10.1021/ACS.JCIM.5B00571>.
- (130) Ferreira, I.; Amarante, T. D.; Weber, G. Salt Dependent Mesoscopic Model for RNA at Multiple Strand Concentrations. *Biophys. Chem.* **2021**, *271*. <https://doi.org/10.1016/J.BPC.2021.106551>.

-
- (131) Song, K. S.; Nimse, S. B.; An, H.; Kim, J.; Nguyen, V. T.; Ta, V. T.; Kim, T. HPV 9G DNAChip: Based on the 9G DNAChip Technology. *J. Virol. Methods* **2012**, *183* (2), 132–138. <https://doi.org/10.1016/J.JVIROMET.2012.04.003>.
- (132) Joda, H.; Beni, V.; Katakis, I.; O'sullivan, C. K. DNA Biosensor Based on Hybridization Refractory Mutation System Approach for Single Mismatch Detection. *Anal. Biochem.* **2015**, *474*, 66–68. <https://doi.org/10.1016/J.AB.2014.11.021>.
- (133) Bahreini, F.; Ramezani, S.; Shahangian, S. S.; Salehi, Z.; Mashayekhi, F. MiR-559 Polymorphism Rs58450758 Is Linked to Breast Cancer. *Br. J. Biomed. Sci.* **2020**, *77* (1), 29–34. <https://doi.org/10.1080/09674845.2019.1683309>.
- (134) Wang, W.; Yang, C.; Nie, H.; Qiu, X.; Zhang, L.; Xiao, Y.; Zhou, W.; Zeng, Q.; Zhang, X.; Wu, Y.; Liu, J.; Ying, M. LIMK2 Acts as an Oncogene in Bladder Cancer and Its Functional SNP in the MicroRNA-135a Binding Site Affects Bladder Cancer Risk. *Int. J. cancer* **2019**, *144* (6), 1345–1355. <https://doi.org/10.1002/IJC.31757>.
- (135) Khan, R.; Shaheen, H.; Mansoor, Q.; Abbasi, S. A.; Fatima, S.; Ammar, A.; Baig, R. M. Genetic Predisposition of SNPs in MiRNA-149 (Rs2292832) and FOXE1 (Rs3758249) in Thyroid Cancer. *Mol. Biol. Rep.* **2021**, *48* (12), 7801–7809. <https://doi.org/10.1007/S11033-021-06795-Y>.
- (136) Zhao, C. N.; Wu, Q.; Mao, Y. M.; Liu, L. N.; Dan, Y. L.; Li, X. M.; Wang, D. G.; Pan, H. F. Elevated Circulating Asymmetric Dimethylarginine Levels in Rheumatoid Arthritis: A Systematic Review and Meta-Analysis. *Amino Acids* **2019**, *51* (5), 773–782. <https://doi.org/10.1007/S00726-019-02714-5>.
- (137) Zhang, N.; Hu, X.; Du, Y.; Du, J. The Role of MiRNAs in Colorectal Cancer Progression and Chemoradiotherapy. *Biomed. Pharmacother.* **2021**, *134*. <https://doi.org/10.1016/J.BIOPHA.2020.111099>.
- (138) Loh, H. Y.; Norman, B. P.; Lai, K. S.; Rahman, N. M. A. N. A.; Alitheen, N. B. M.; Osman, M. A. The Regulatory Role of MicroRNAs in Breast Cancer. *Int. J. Mol. Sci.* **2019**, *20* (19). <https://doi.org/10.3390/IJMS20194940>.
- (139) Vanacore, D.; Boccellino, M.; Rossetti, S.; Cavaliere, C.; D'Aniello, C.; Di Franco, R.; Romano, F. J.; Montanari, M.; La Mantia, E.; Piscitelli, R.; Nocerino, F.; Cappuccio, F.; Grimaldi, G.; Izzo, A.; Castaldo, L.; Pepe, M. F.; Malzone, M. G.; Iovane, G.; Ametrano, G.; Stiuso, P.; Quagliuolo, L.; Barberio, D.; Perdonà, S.; Muto, P.; Montella, M.; Maiolino, P.; Veneziani, B. M.; Botti, G.; Caraglia, M.; Facchini, G. Micrnas in Prostate Cancer: An Overview. *Oncotarget* **2017**, *8* (30), 50240–50251. <https://doi.org/10.18632/ONCOTARGET.16933>.
- (140) Zhong, S.; Golpon, H.; Zardo, P.; Borlak, J. MiRNAs in Lung Cancer. A Systematic Review Identifies Predictive and Prognostic MiRNA Candidates for Precision Medicine in Lung Cancer. *Transl. Res.* **2021**, *230*, 164–196. <https://doi.org/10.1016/J.TRSL.2020.11.012>.
- (141) Mao, Y.; Xue, P.; Li, L.; Xu, P.; Cai, Y.; Chu, X.; Jiang, P.; Zhu, S. Bioinformatics Analysis of
-

- MRNA and MiRNA Microarray to Identify the Key MiRNA-gene Pairs in Small-cell Lung Cancer. *Mol. Med. Rep.* **2019**, 20 (3), 2199–2208. <https://doi.org/10.3892/MMR.2019.10441>.
- (142) Li, Z.; Wang, Y.; Zhu, Y. Association of MiRNA-146a Rs2910164 and MiRNA-196 Rs11614913 Polymorphisms in Patients with Ulcerative Colitis: A Meta-Analysis and Review. *Medicine (Baltimore)*. **2018**, 97 (39). <https://doi.org/10.1097/MD.00000000000012294>.
- (143) Benderska, N.; Dittrich, A. L.; Knaup, S.; Rau, T. T.; Neufert, C.; Wach, S.; Fahlbusch, F. B.; Rauh, M.; Wirtz, R. M.; Agaimy, A.; Srinivasan, S.; Mahadevan, V.; Rümmele, P.; Rapti, E.; Gazouli, M.; Hartmann, A.; Schneider-Stock, R. MiRNA-26b Overexpression in Ulcerative Colitis-Associated Carcinogenesis. *Inflamm. Bowel Dis.* **2015**, 21 (9), 2039–2051. <https://doi.org/10.1097/MIB.0000000000000453>.
- (144) Su, X.; Hu, Y.; Li, Y.; Cao, J. L.; Wang, X. Q.; Ma, X.; Xia, H. F. The Polymorphism of Rs6505162 in the MIR423 Coding Region and Recurrent Pregnancy Loss. *Reproduction* **2015**, 150 (1), 65–76. <https://doi.org/10.1530/REP-15-0007>.
- (145) Li, M.; Fu, W.; Wo, L.; Shu, X.; Liu, F.; Li, C. MiR-128 and Its Target Genes in Tumorigenesis and Metastasis. *Exp. Cell Res.* **2013**, 319 (20), 3059–3064. <https://doi.org/10.1016/J.YEXCR.2013.07.031>.
- (146) Zhuang, L.; Xu, L.; Wang, P.; Meng, Z. Serum MiR-128-2 Serves as a Prognostic Marker for Patients with Hepatocellular Carcinoma. *PLoS One* **2015**, 10 (2). <https://doi.org/10.1371/JOURNAL.PONE.0117274>.
- (147) Donzelli, S.; Fontemaggi, G.; Fazi, F.; Di Agostino, S.; Padula, F.; Biagioni, F.; Muti, P.; Strano, S.; Blandino, G. MicroRNA-128-2 Targets the Transcriptional Repressor E2F5 Enhancing Mutant P53 Gain of Function. *Cell Death Differ.* **2012**, 19 (6), 1038–1048. <https://doi.org/10.1038/CDD.2011.190>.
- (148) Weber, G.; Haslam, N.; Whiteford, N.; Prügel-Bennett, A.; Essex, J. W.; Neylon, C. Thermal Equivalence of DNA Duplexes without Calculation of Melting Temperature. *Nat. Phys.* **2006** 21 **2005**, 2 (1), 55–59. <https://doi.org/10.1038/nphys189>.
- (149) Kalies, S.; Kuetemeyer, K.; Heisterkamp, A. Mechanisms of High-Order Photobleaching and Its Relationship to Intracellular Ablation. *Biomed. Opt. Express* **2011**, 2 (4), 805. <https://doi.org/10.1364/BOE.2.000816>.
- (150) Metcalf, G. A. D.; Shibakawa, A.; Patel, H.; Sita-Lumsden, A.; Zivi, A.; Rama, N.; Bevan, C. L.; Ladame, S. Amplification-Free Detection of Circulating MicroRNA Biomarkers from Body Fluids Based on Fluorogenic Oligonucleotide-Templated Reaction between Engineered Peptide Nucleic Acid Probes: Application to Prostate Cancer Diagnosis. *Anal. Chem.* **2016**, 88 (16), 8091–8098. <https://doi.org/10.1021/ACS.ANALCHEM.6B01594>.
- (151) Xia, J.; Xu, T.; Qing, J.; Wang, L.; Tang, J. Detection of Single Nucleotide Polymorphisms by Fluorescence Embedded Dye SYBR Green I Based on Graphene Oxide. *Front. Chem.* **2021**, 9. <https://doi.org/10.3389/FCHEM.2021.631959>.
- (152) Nachman, M. W.; Crowell, S. L. Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics* **2000**, 156 (1), 297–304. <https://doi.org/10.1093/GENETICS/156.1.297>.

-
- (153) Xue, Y.; Wang, Q.; Long, Q.; Ng, B. L.; Swerdlow, H.; Burton, J.; Skuce, C.; Taylor, R.; Abdellah, Z.; Zhao, Y.; MacArthur, D. G.; Quail, M. A.; Carter, N. P.; Yang, H.; Tyler-Smith, C. Human Y Chromosome Base-Substitution Mutation Rate Measured by Direct Sequencing in a Deep-Rooting Pedigree. *Curr. Biol.* **2009**, *19* (17), 1453–1457. <https://doi.org/10.1016/J.CUB.2009.07.032>.
- (154) Chang, F.; Sun, Y.; Yang, D.; Yang, W.; Sun, Y.; Liu, C.; Li, Z. Specific Detection of RNA Mutation at Single-Base Resolution by Coupling the Isothermal Exponential Amplification Reaction (EXPAR) with Chimeric DNA Probe-Aided Precise RNA Disconnection at the Mutation Site. *Chem. Commun. (Camb.)* **2019**, *55* (48), 6934–6937. <https://doi.org/10.1039/C9CC02700A>.
- (155) Kim, D. M.; Seo, J.; Kim, D. W.; Jeong, W.; Hwang, S. H.; Kim, D. E. Fluorometric Detection of Single-Nucleotide Mutations Using Tandem Gene Amplification. *Sensors Actuators B Chem.* **2020**, *314*, 128071. <https://doi.org/10.1016/J.SNB.2020.128071>.
- (156) Weber, G. Mesoscopic Model Parametrization of Hydrogen Bonds and Stacking Interactions of RNA from Melting Temperatures. *Nucleic Acids Res.* **2013**, *41* (1). <https://doi.org/10.1093/NAR/GKS964>.
- (157) W. H. Press; B. P. Flannery; S. A. Teukolsky; W. T. Vetterling. Numerical Recipes in C—the Art of Scientific Computing, . *Math. Gaz.* **1989**, *73* (464), 167–170. <https://doi.org/10.2307/3619708>.
- (158) Kawasaki, A. M.; Casper, M. D.; Freier, S. M.; Lesnik, E. A.; Zounes, M. C.; Cummins, L. L.; Gonzalez, C.; Dan Cook, P. Uniformly Modified 2'-Deoxy-2'-Fluoro Phosphorothioate Oligonucleotides as Nuclease-Resistant Antisense Compounds with High Affinity and Specificity for RNA Targets. *J. Med. Chem.* **1993**, *36* (7), 831–841. <https://doi.org/10.1021/JM00059A007>.
- (159) Nakano, S. I.; Fujimoto, M.; Hara, H.; Sugimoto, N. Nucleic Acid Duplex Stability: Influence of Base Composition on Cation Effects. *Nucleic Acids Res.* **1999**, *27* (14), 2957–2965. <https://doi.org/10.1093/NAR/27.14.2957>.
- (160) Lesnik, E. A.; Freier, S. M. Relative Thermodynamic Stability of DNA, RNA, and DNA:RNA Hybrid Duplexes: Relationship with Base Composition and Structure. *Biochemistry* **1995**, *34* (34), 10807–10815. <https://doi.org/10.1021/BI00034A013>.
- (161) Gyi, J. I.; Conn, G. L.; Lane, A. N.; Brown, T. Comparison of the Thermodynamic Stabilities and Solution Conformations of DNA:RNA Hybrids Containing Purine-Rich and Pyrimidine-Rich Strands with DNA and RNA Duplexes. *Biochemistry* **1996**, *35* (38), 12538–12548. <https://doi.org/10.1021/BI960948Z>.
- (162) Banerjee, D.; Tateishi-Karimata, H.; Ohyama, T.; Ghosh, S.; Endoh, T.; Takahashi, S.; Sugimoto, N. Improved Nearest-Neighbor Parameters for the Stability of RNA/DNA Hybrids under a Physiological Condition. *Nucleic Acids Res.* **2020**, *48* (21), 12042–12054. <https://doi.org/10.1093/NAR/GKAA572>.
-

- (163) Caruthers, M. H. Gene Synthesis Machines: DNA Chemistry and Its Uses. *Science* **1985**, 230 (4723), 281–285. <https://doi.org/10.1126/SCIENCE.3863253>.
- (164) Bradley, C. A. First Antisense Drug Is Approved with Fleeting Success. *Nat. Res.* 2021 **2019**.
- (165) Mullard, A. FDA Approves Fifth RNAi Drug — Alnylam’s next-Gen HATTR Treatment. *Nat. Rev. Drug Discov.* **2022**, 21 (8), 548–549. <https://doi.org/10.1038/D41573-022-00118-X>.
- (166) Pon, R. T.; Yu, S.; Sanghvi, Y. S. Tandem Oligonucleotide Synthesis on Solid-Phase Supports for the Production of Multiple Oligonucleotides. *J. Org. Chem.* **2002**, 67 (3), 856–864. <https://doi.org/10.1021/jo0160773>.
- (167) Hardy, P. M.; Holland, D.; Scott, S.; Garman, A. J.; Newton, C. R.; Mclean, M. J. Reagents for the RepARATION of Two Oligonucleotides per Synthesis (TO PST. *Nucleic Acids Res.* **1994**, 22 (15), 2998–3004.
- (168) Yamamoto, K.; Fuchi, Y.; Okabe, M.; Osawa, T.; Ito, Y.; Hari, Y. New Cleavable Spacers for Tandem Synthesis of Multiple Oligo-nucleotides. *Synth.* **2021**, 53 (23), 4440–4448. <https://doi.org/10.1055/A-1538-9883>.
- (169) Arar, K. *Methods and compositions for the tandem synthesis of two or more oligonucleotides on the same solid support* - US7615629B2. Patent. <https://patents.google.com/patent/US7615629B2/en> (accessed 2022-11-25).
- (170) Hervé, B.; André, R.; Robert, T. Phosphoramidite Reagents for the Easy Preparation of Polylabelled Oligonucleotide Probes. *Nucleosides and Nucleotides* **1991**, 10 (1–3), 363–366. <https://doi.org/10.1080/07328319108046480>.
- (171) Xu, J.; Duffy, C. D.; Chan, C. K. W.; Sutherland, J. D. Solid-Phase Synthesis and Hybridization Behavior of Partially 2’/3’-O-Acetylated RNA Oligonucleotides. *J. Org. Chem.* **2014**, 79 (8), 3311–3326. https://doi.org/10.1021/JO5002824/SUPPL_FILE/JO5002824_SI_002.PDF.
- (172) Reese, C. B.; Trentham, D. R. Acyl Migration in Ribonucleoside Derivatives. *Tetrahedron Lett.* **1965**, 6 (29), 2467–2472. [https://doi.org/10.1016/S0040-4039\(01\)84008-9](https://doi.org/10.1016/S0040-4039(01)84008-9).
- (173) Kempe, T.; Chow, F.; Sundquist, W. I.; Nardi, T. J.; Paulson, B.; Peterson, S. M. Selective 2’-Benzoylation at the Cis 2’,3’-Diols of Protected Ribonucleosides. New Solid Phase Synthesis of RNA and DNA-RNA Mixtures. *Nucleic Acids Res.* **1982**, 10 (21), 6695–6714. <https://doi.org/10.1093/NAR/10.21.6695>.
- (174) Shen, S. J.; Song, Y.; Ren, X. Y.; Xu, Y. L.; Zhou, Y. D.; Liang, Z. Y.; Sun, Q. MicroRNA-27b-3p Promotes Tumor Progression and Metastasis by Inhibiting Peroxisome Proliferator-Activated Receptor Gamma in Triple-Negative Breast Cancer. *Front. Oncol.* **2020**, 10, 1371. <https://doi.org/10.3389/FONC.2020.01371/BIBTEX>.
- (175) Zhou, J.; Tian, Y.; Li, J.; Lu, B.; Sun, M.; Zou, Y.; Kong, R.; Luo, Y.; Shi, Y.; Wang, K.; Ji, G. MiR-206 Is down-Regulated in Breast Cancer and Inhibits Cell Proliferation through the up-Regulation of CyclinD2. *Biochem. Biophys. Res. Commun.* **2013**, 433 (2), 207–212. <https://doi.org/10.1016/j.bbrc.2013.02.084>.
- (176) Meng, R.; Fang, J.; Yu, Y.; Hou, L. K.; Chi, J. R.; Chen, A. X.; Zhao, Y.; Cao, X. C. MiR-129-

-
- 5p Suppresses Breast Cancer Proliferation by Targeting CBX4. *Neoplasma* **2018**, 65 (4), 572–578. https://doi.org/10.4149/NEO_2018_170814N530.
- (177) Stepicheva, N. A.; Song, J. L. Function and Regulation of MicroRNA-31 in Development and Disease. *Mol. Reprod. Dev.* **2016**, 83 (8), 654–674. <https://doi.org/10.1002/MRD.22678>.
- (178) Gandham, S. K.; Rao, M.; Shah, A.; Trivedi, M. S.; Amiji, M. M. Combination MicroRNA-Based Cellular Reprogramming with Paclitaxel Enhances Therapeutic Efficacy in a Relapsed and Multidrug-Resistant Model of Epithelial Ovarian Cancer. **2022**. <https://doi.org/10.1016/j.omto.2022.03.005>.
- (179) Wu, J.; Huang, J.; Kuang, S.; Chen, J.; Li, X.; Chen, B.; Wang, J.; Cheng, D.; Shuai, X. Synergistic MicroRNA Therapy in Liver Fibrotic Rat Using MRI-Visible Nanocarrier Targeting Hepatic Stellate Cells. *Adv. Sci. (Weinheim, Baden-Wurttemberg, Ger.)* **2019**, 6 (5). <https://doi.org/10.1002/ADVS.201801809>.
- (180) *miRBase: the microRNA database*. <https://www.mirbase.org/> (accessed 2022-12-01).
- (181) Debarge, S.; Balzarini, J.; Maguire, A. R. Design and Synthesis of α -Carboxy Phosphononucleosides. *J. Org. Chem* **2011**, 76, 105. <https://doi.org/10.1021/jo101738e>.
- (182) Lyttle, M. H.; Dick, D. J.; Hudson, D.; Cook, R. M. A Phosphate Bound Universal Linker for DNA Synthesis. *Nucleosides and Nucleotides* **1999**, 18 (8), 1809–1824. <https://doi.org/10.1080/07328319908044845>.
- (183) Shen, S. J.; Song, Y.; Ren, X. Y.; Xu, Y. L.; Zhou, Y. D.; Liang, Z. Y.; Sun, Q. MicroRNA-27b-3p Promotes Tumor Progression and Metastasis by Inhibiting Peroxisome Proliferator-Activated Receptor Gamma in Triple-Negative Breast Cancer. *Front. Oncol.* **2020**, 10. <https://doi.org/10.3389/FONC.2020.01371>.
- (184) Albert, T. J.; Molla, M. N.; Muzny, D. M.; Nazareth, L.; Wheeler, D.; Song, X.; Richmond, T. A.; Middle, C. M.; Rodesch, M. J.; Packard, C. J.; Weinstock, G. M.; Gibbs, R. A. Direct Selection of Human Genomic Loci by Microarray Hybridization. *Nat. Methods* **2007**, 4 (11), 903–905. <https://doi.org/10.1038/NMETH1111>.
- (185) Hodges, E.; Xuan, Z.; Balija, V.; Kramer, M.; Molla, M. N.; Smith, S. W.; Middle, C. M.; Rodesch, M. J.; Albert, T. J.; Hannon, G. J.; McCombie, W. R. Genome-Wide in Situ Exon Capture for Selective Resequencing. *Nat. Genet.* **2007**, 39 (12), 1522–1527. <https://doi.org/10.1038/NG.2007.42>.
- (186) Porreca, G. J.; Zhang, K.; Li, J. B.; Xie, B.; Austin, D.; Vassallo, S. L.; LeProust, E. M.; Peck, B. J.; Emig, C. J.; Dahl, F.; Gao, Y.; Church, G. M.; Shendure, J. Multiplex Amplification of Large Sets of Human Exons. *Nat. Methods* **2007**, 4 (11), 931–936. <https://doi.org/10.1038/NMETH1110>.
- (187) Myllykangas, S.; Buenrostro, J. D.; Natsoulis, G.; Bell, J. M.; Ji, H. P. Efficient Targeted Resequencing of Human Germline and Cancer Genomes by Oligonucleotide-Selective Sequencing. *Nat. Biotechnol.* **2011**, 29 (11), 1024–1027. <https://doi.org/10.1038/NBT.1996>.
-

- (188) Hopmans, E. S.; Natsoulis, G.; Bell, J. M.; Grimes, S. M.; Sieh, W.; Ji, H. P. A Programmable Method for Massively Parallel Targeted Sequencing. *Nucleic Acids Res.* **2014**, *42* (10). <https://doi.org/10.1093/NAR/GKU282>.
- (189) Lau, B. T.; Ji, H. P. Covalent “Click Chemistry”-Based Attachment of DNA onto Solid Phase Enables Iterative Molecular Analysis. *Anal. Chem.* **2019**, *91* (3), 1706–1710. https://doi.org/10.1021/ACS.ANALCHEM.8B05139/ASSET/IMAGES/LARGE/AC-2018-05139Z_0005.JPEG.
- (190) Carboni, R. A.; Lindsey, R. V. Reactions of Tetrazines with Unsaturated Compounds. A New Synthesis of Pyridazines. *J. Am. Chem. Soc.* **1959**, *81* (16), 4342–4346. https://doi.org/10.1021/JA01525A060/ASSET/JA01525A060.FP.PNG_V03.
- (191) Oliveira, B. L.; Guo, Z.; Bernardes, G. J. L. Inverse Electron Demand Diels–Alder Reactions in Chemical Biology. *Chem. Soc. Rev.* **2017**, *46* (16), 4895–4950. <https://doi.org/10.1039/C7CS00184C>.
- (192) Shin, G. W.; Greer, S. U.; Hopmans, E.; Grimes, S. M.; Lee, H. J.; Zhao, L.; Miotke, L.; Suarez, C.; Almeda, A. F.; Haraldsdottir, S.; Ji, H. P. Profiling Diverse Sequence Tandem Repeats in Colorectal Cancer Reveals Co-Occurrence of Microsatellite and Chromosomal Instability Involving Chromosome 8. *Genome Med.* **2021**, *13* (1), 1–18. <https://doi.org/10.1186/S13073-021-00958-Z/FIGURES/7>.
- (193) Hölz, K.; Pavlic, A.; Lietard, J.; Somoza, M. M. Specificity and Efficiency of the Uracil DNA Glycosylase-Mediated Strand Cleavage Surveyed on Large Sequence Libraries. *Sci. Reports* **2019**, *9* (1), 1–12. <https://doi.org/10.1038/s41598-019-54044-x>.
- (194) Garcia-Perez, L.; van Eggermond, M. C. J. A.; Maietta, E.; van der Hoorn, M. L. P.; Pike-Overzet, K.; Staal, F. J. T. A Novel Branched DNA-Based Flowcytometric Method for Single-Cell Characterization of Gene Therapy Products and Expression of Therapeutic Genes. *Front. Immunol.* **2021**, *11*, 3533. <https://doi.org/10.3389/FIMMU.2020.607991/BIBTEX>.
- (195) NHS ester labeling of amino biomolecules. <https://www.lumiprobe.com/protocols/nhs-ester-labeling> (accessed 2023-01-19).
- (196) DNA Polymerase Strand Displacement Activity | NEB. <https://international.neb.com/products/pcr-qpcr-and-amplification-technologies/pcr-qpcr-and-amplification-technologies/dna-polymerase-strand-displacement-activity> (accessed 2023-01-19).
- (197) Illumina. NextSeq 500 and NextSeq 550 Sequencing Systems Denature and Dilute Libraries Guide (15048776). **2022**.

Appendix

A1 List of Publications and Conferences

Publications

Main Publications

1. **Slott, S.** & Astakhova, K. (2022). MicroRNA Pools Synthesized Using Tandem Solid-Phase Oligonucleotide Synthesis. (Submitted to JOC)
2. **Slott, S.**, Krüger-Jensen, C. S., Ferreira, I., Pedersen, N. B. & Astakhova, K. (2022). Mutations in microRNA-128-2-3p identified with amplification-free hybridization assay. (Submitted to OBC)
3. Ferreira, I., **Slott, S.**, Astakhova, K., & Weber, G. (2021). Complete Mesoscopic Parameterization of Single LNA Modifications in DNA Applied to Oncogene Probe Design. *Journal of chemical information and modeling*, 61(7), 3615–3624.

Other Publications

4. Muniz, M. I., Bustos, A. H., **Slott, S.**, Astakhova, K. & Weber, G. (2022) Cation valence dependence of hydrogen bond and stacking potentials in DNA mesoscopic models, *Biophysical Chemistry*.
5. Fajardo, P., Taskova, M., Martín-Serrano, M. A., Hansen, J., **Slott, S.**, Jakobsen, A. K., Wibom, M. L., Salegi, B., Muñoz, A., Barbachano, A., Sharma, A., Gubatan, J. M., Habtezion, A., Sanz-Ezquerro, J. J., Astakhova, K. & Cuenda, A. (2022). p38 γ and p38 δ as biomarkers in the interplay of colon cancer and inflammatory bowel diseases. *Cancer communications (London, England)*, 10.1002/cac2.12331.
6. van Hees, M., **Slott, S.**, Hansen, A. H., Kim, H. S., Ji, H. P., & Astakhova, K. (2022). New approaches to moderate CRISPR-Cas9 activity: Addressing issues of cellular uptake and endosomal escape. *Molecular therapy : the journal of the American Society of Gene Therapy*, 30(1), 32–46.
7. Kofoed Andersen, C., Khatri, S., Hansen, J., **Slott, S.**, Pavan Parvathaneni, R., Mendes, A. C., Chronakis, I. S., Hung, S. C., Rajasekaran, N., Ma, Z., Zhu, S., Dai, H., Mellins, E. D., & Astakhova, K. (2021). Carbon Nanotubes-Potent Carriers for Targeted Drug Delivery in Rheumatoid Arthritis. *Pharmaceutics*, 13(4), 453.

8. Ray, R. M., Hansen, A. H., **Slott, S.**, Taskova, M., Astakhova, K., & Morris, K. V. (2019). Control of LDL Uptake in Human Cells by Targeting the LDLR Regulatory Long Non-coding RNA BM450697. *Molecular therapy. Nucleic acids*, 17, 264–276.
9. Kjelsen, C., **Slott, S.**, Elverdal, P. L., Sheppard, C. L., Kapatai, G., Fry, N. K., ... Duus, J. Ø. (2018). Discovery and description of a new serogroup 7 *Streptococcus pneumoniae* serotype, 7D, and structural analysis of 7C and 7D. *Carbohydrate Research*, 463, 24–31.

Conferences

2022

1. PhD symposium DTU Chemistry - November 2022 (Poster)
2. International Round Table on Nucleosides, Nucleotides and Nucleic Acids – Stockholm, August 2022 (Poster)

A2 Co-Author Statements



PhD Administration & Development

Declaration of co-authorship at DTU

If a PhD thesis contains articles (i.e. published journal and conference articles, unpublished manuscripts, chapters, etc.) written in collaboration with other researchers, a co-author statement verifying the PhD student's contribution to each article should be made.

If an article is written in collaboration with three or less researchers (including the PhD student), all researchers must sign the statement. However, if an article has more than three authors the statement may be signed by a representative sample, cf. article 12, section 4 and 5 of the Ministerial Order No. 1039, 27 August 2013. A representative sample consists of minimum three authors, which is comprised of the first author, the corresponding author, the senior author, and 1-2 authors (preferably international/non-supervisor authors).

DTU has implemented the Danish Code of Conduct for Research Integrity, which states the following regarding attribution of authorship:

"Attribution of authorship should in general be based on criteria a-d adopted from the Vancouver guidelines¹, and all individuals who meet these criteria should be recognized as authors:

- a. Substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work, *and*
- b. drafting the work or revising it critically for important intellectual content, *and*
- c. final approval of the version to be published, *and*
- d. agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved."²

For more information regarding definition of co-authorship and examples of authorship conflicts, we refer to DTU Code of Conduct for Research Integrity (pp. 19-22).


By signing this declaration, co-authors permit the PhD student to reuse whole or parts of co-authored articles in their PhD thesis, under the condition that co-authors are acknowledged in connection with the reused text or figure.

It is **important** to note that it is the responsibility of the PhD student to obtain permission from the publisher to use the article in the PhD thesis³



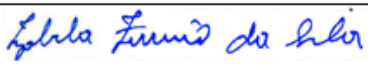

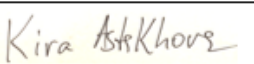
¹International Committee of Medical Journal Editors – Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals, updated December 2016

²DTU Code of Conduct for Research Integrity (E-book p. 19)

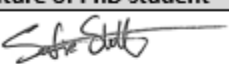
³Many journals will allow you to use only the post-print version of your article, meaning the accepted version of the article, without the publisher's final formatting. In the event that your article is submitted, but still under review, you should of course use the latest submitted version of your article in your thesis. Always remember to check your publisher's guidelines on reuse

Title of article		
Mutations in microRNA-128-2-3p identified with amplification-free hybridization assay		
Journal/conference		
Organic and Biomolecular Chemistry		
Author(s)		
Sofie Slott, Cecilie Krüger-Jensen, Izabela Ferreira, Nadia Bom Pedersen, Kira Astakhova		
Name (capital letters) and signature of PhD student		
SOFIE SLOTT 		
PhD student's date of birth		
09/07/1992		
Declaration of the PhD student's contribution		
<i>For each category in the table below, please specify the PhD student's contribution to the article as appropriate (please do not fill in with names or x's)</i>		
Category	Minor contribution to the work <i>(please specify the nature of the PhD student's contribution)</i>	Substantial contribution to the work <i>(please specify the nature of the PhD student's contribution)</i>
Formulation of the conceptual framework and/or planning of the design of the study including scientific questions	Participating in discussions of all sections of the paper.	Literature study and conceived the idea with supervisor
Carrying out of experiments/data collection and analysis/interpretation of results		Designed of the assay and conducted the experimental procedure. Conducted data analysis.
Writing of the article/revising the manuscript for intellectual content		Wrote initial draft of manuscript and revised final manuscript
Signatures		


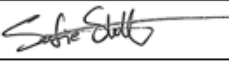
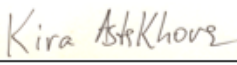
of published articles. Most journals, unless open access, have an embargo period on published articles, meaning that within this period you cannot freely use the article. Check your publisher's rules on this issue.

Title of article			
Mutations in microRNA-128-2-3p identified with amplification-free hybridization assay			
Journal/conference			
Organic and Biomolecular Chemistry			
Author(s)			
Sofie Slott, Cecilie Krüger-Jensen, Izabela Ferreira, Nadia Bom Pedersen, Kira Astakhova			
Name (capital letters) and signature of PhD student			
SOFIE SLOTT 			
PhD student's date of birth			
09/07/1992			
Date	Name	Title	Signature
18/01/2023	Sofie Slott	PhD Student	
19/01/2023	Izabela Ferreira	PhD	
18/01/2023	Nadia Bom Pedersen	Research assistant	
23.01.2023	Kira Astakhova	Associate Professor	

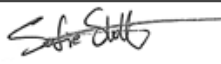
Please note that by signing this declaration, co-authors permit the PhD student to reuse whole or parts of co-authored articles in their PhD thesis, under the condition that co-authors are acknowledged in connection with the reused text or figure.

Title of article		
MicroRNA Pools Synthesized Using Tandem Solid-Phase Oligonucleotide Synthesis		
Journal/conference		
The Journal of Organic Chemistry		
Author(s)		
Sofie Slott, Kira Astakhova		
Name (capital letters) and signature of PhD student		
SOFIE SLOTT 		
PhD student's date of birth		
09/07/1992		
Declaration of the PhD student's contribution		
For each category in the table below, please specify the PhD student's contribution to the article as appropriate (please do not fill in with names or x's)		
Category	Minor contribution to the work (please specify the nature of the PhD student's contribution)	Substantial contribution to the work (please specify the nature of the PhD student's contribution)
Formulation of the conceptual framework and/or planning of the design of the study including scientific questions		Literature study. Conceived idea with supervisor. Prepared the paper outline.
Carrying out of experiments/data collection and analysis/interpretation of results		Designed and performed the experiments. Conducted data collection and analysis
Writing of the article/revising the manuscript for intellectual content		Wrote the manuscript. Applied comments and correction
Signatures		

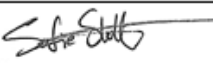
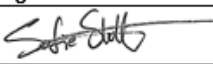
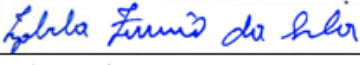
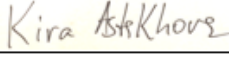

of published articles. Most journals, unless open access, have an embargo period on published articles, meaning that within this period you cannot freely use the article. Check your publisher's rules on this issue.

Title of article			
MicroRNA Pools Synthesized Using Tandem Solid-Phase Oligonucleotide Synthesis			
Journal/conference			
The Journal of Organic Chemistry			
Author(s)			
Sofie Slott, Kira Astakhova			
Name (capital letters) and signature of PhD student			
SOFIE SLOTT 			
PhD student's date of birth			
09/07/1992			
Date	Name	Title	Signature
18/01/2023	Sofie Slott	PhD Student	
23.01.2023	Kira Astakhova	Associate Professor	

Please note that by signing this declaration, co-authors permit the PhD student to reuse whole or parts of co-authored articles in their PhD thesis, under the condition that co-authors are acknowledged in connection with the reused text or figure.

Title of article		
Complete Mesoscopic Parameterization of Single LNA Modifications in DNA Applied to Oncogene Probe Design		
Journal/conference		
Journal of Chemical Information and Modeling		
Author(s)		
Izabela Ferreira, Sofie Slott, Kira Astakhova, and Gerald Weber		
Name (capital letters) and signature of PhD student		
SOFIE SLOTT 		
PhD student's date of birth		
09/07/1992		
Declaration of the PhD student's contribution		
<i>For each category in the table below, please specify the PhD student's contribution to the article as appropriate (please do not fill in with names or x's)</i>		
Category	Minor contribution to the work <i>(please specify the nature of the PhD student's contribution)</i>	Substantial contribution to the work <i>(please specify the nature of the PhD student's contribution)</i>
Formulation of the conceptual framework and/or planning of the design of the study including scientific questions	Discussion of scientific scope and goal of the paper	Participated in planning experimental design of study including scientific discussion
Carrying out of experiments/data collection and analysis/interpretation of results		Performed oligonucleotide synthesis. Conducted data collection and analysis.
Writing of the article/revising the manuscript for intellectual content	Proof reading of the manuscript.	Wrote part related to oligonucleotide synthesis
Signatures		

of published articles. Most journals, unless open access, have an embargo period on published articles, meaning that within this period you cannot freely use the article. Check your publisher's rules on this issue.

Title of article			
Complete Mesoscopic Parameterization of Single LNA Modifications in DNA Applied to Oncogene Probe Design			
Journal/conference			
Journal of Chemical Information and Modeling			
Author(s)			
Izabela Ferreira, Sofie Slott, Kira Astakhova, and Gerald Weber			
Name (capital letters) and signature of PhD student			
SOFIE SLOTT 			
PhD student's date of birth			
09/07/1992			
Date	Name	Title	Signature
18/01/2023	Sofie Slott	PhD Student	
19/01/2023	Izabela Ferreira	PhD	
23.1.2023	Kira Astakhova	Associate Professor	
20/01/2023	Gerald Weber	Professor	

Please note that by signing this declaration, co-authors permit the PhD student to reuse whole or parts of co-authored articles in their PhD thesis, under the condition that co-authors are acknowledged in connection with the reused text or figure.

A3 Publication: Complete Mesoscopic Parameterization of Single LNA Modifications in DNA Applied to Oncogene Probe Design

Complete Mesoscopic Parameterization of Single LNA Modifications in DNA Applied to Oncogene Probe Design

Izabela Ferreira, Sofie Slott, Kira Astakhova,* and Gerald Weber*

Cite This: *J. Chem. Inf. Model.* 2021, 61, 3615–3624

Read Online

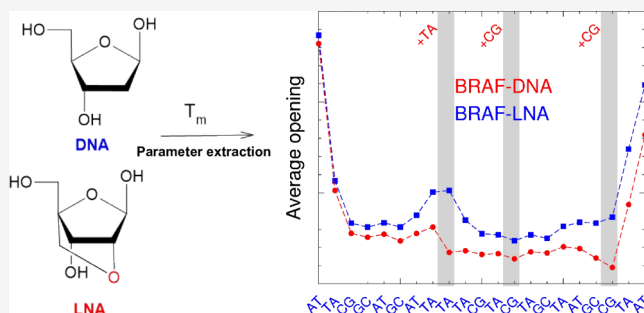
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The use of mesoscopic models to describe the thermodynamic properties of locked nucleic acid (LNA)-modified nucleotides can provide useful insights into their properties, such as hydrogen-bonding and stacking interactions. In addition, the mesoscopic parameters can be used to optimize LNA insertion in probes, to achieve accurate melting temperature predictions, and to obtain duplex opening profiles at the base-pair level. Here, we applied this type of model to parameterize a large set of melting temperatures for LNA-modified sequences, from published sources, covering all possible nearest-neighbor configurations. We have found a very large increase in Morse potentials, which indicates very strong hydrogen bonding as the main cause of improved LNA thermodynamic stability. LNA-modified adenine–thymine (AT) was found to have similar hydrogen bonding to unmodified cytosine–guanine (CG) base pairs, while for LNA CG, we found exceptionally large hydrogen bonding. In contrast, stacking interactions, which were thought to be behind the stability of LNA, were similar to unmodified DNA in most cases. We applied the new LNA parameters to the design of *BRAF*, *KRAS*, and *EGFR* oncogene variants by testing all possible LNA modifications. Selected sequences were then synthesized and had their hybridization temperatures measured, achieving a prediction accuracy within 1 °C. We performed a detailed base-pair opening analysis to discuss specific aspects of these probe hybridizations that may be relevant for probe design.



INTRODUCTION

Locked nucleic acid, or LNA, are DNA/RNA analogues with a modification that adds a methylene bridge between the 2'-oxygen and 4'-carbon of the ribose sugar, locking it in a C3'-endo/N conformation reducing its flexibility, resembling an RNA helix. They were synthesized in the late 1990s by several groups.^{1–6} One of the most remarkable properties of LNA is the ability to increase the stability of chimera duplexes in both DNA and RNA. This stability has a direct impact on the denaturation temperatures of the sequences. Several studies have shown a considerable increase in the melting temperature^{7–10} reaching up to 10° C per added modification. The presence of three LNAs at the 5' and 3' ends is enough to increase the half-life of the nucleotide,¹¹ and LNA-modified probes can selectively capture genomic DNA sequences.¹²

Its base pairing specificity and mismatch sensitivity make it attractive for use in many applications, such as polymerase chain reaction (PCR) detection and diagnosis,^{13–15} high binding diagnostic probes,^{16–18} stability improvement and hybridization efficiency in loop-mediated isothermal amplification (LAMP) detection,¹⁹ improvement of targeting, specificity and stability in antisense oligonucleotides (ASOs),^{20–22} aptamers,^{23–25} and siRNA approaches.^{26,27} LNA is also used in DNazymes and LNazymes to improve targeting and cleavage efficiency,^{28–30} molecular beacons,³¹

and as enhancers to RNA *in situ* hybridization^{32,33} and as direct antagonist in miRNA silencing.^{34–36}

One application of LNA that is of particular interest is to detect cancer mutations directly in DNA present in patients' blood samples (so-called circulating tumor DNA, or ctDNA) and in cellular mRNA.³⁷ Recently, it was shown that mutations in oncogene regions of ctDNA are attractive biomarkers for early diagnostics and monitoring of different cancer diseases.³⁸ Thus, the ability to effectively detect these oncogenes in ctDNA would open up new opportunities for point-of-care diagnostics and treatment of cancer. However, ctDNAs are typically present in the very complex blood medium at extremely low concentrations. This makes ctDNAs challenging for detection and require ultrasensitive and specific probes.³⁷ In turn, mRNA detection using new probes could provide valuable insights into cancer development.³⁹

Received: April 27, 2021

Published: July 12, 2021



The design of ultrasensitive and specific probes with LNA modifications requires a good understanding of intramolecular interactions, specifically hydrogen bonding and stacking, that gives LNA its increased thermodynamic stability. Yet, despite the intensive use of LNA over the last two decades, the origin of its enhanced base pairing stability is still not sufficiently understood, and some conflicting explanations have been given. It was suggested by thermodynamic studies that a decrease in the entropy of duplex formation and improved stacking on the duplex may both play a part, and there may be a negative contribution in enthalpy due to the disruption of the hydrogen bonds.^{40,41} Additionally, an association between a favorable enthalpy increase and a more pronounced stacking interaction was proposed; however, this is context-dependent and might be influenced by flanking base pairs.¹⁰ Some studies suggested that the conformation change induces a reduction in entropy, being localized at the level of individual base pairs, consequently increasing the overall thermodynamic helix stability.^{42,43} Several studies attempted to address the source of the stabilizing effect provided by LNA incorporation from a structural point of view. For instance, NMR and X-ray diffraction studies accounted for the stabilization in a stacking enhancement induced by the modification.^{44–47} However, another NMR assay ruled out both stacking interaction or hydrogen bonding as a source of the stabilization effect, calling for more detailed studies with hybridized water, that is, the water molecule bridging two hydrogen bonds simultaneously.⁴⁸

Nearest-neighbor (NN) models were extensively used to predict the thermodynamic properties for LNA-modified probes.^{41,43,49,50} However, these types of models provide little insight into the intramolecular interactions and were unable to indicate the physical source of the increased LNA stability. On the other hand, mesoscopic models make use of the same melting temperature data as NN models, but have the potential to provide insights into the hydrogen bonds and stacking interactions.⁵¹ We have shown that it is possible to obtain a good insight into these interactions from melting temperatures,⁵¹ and we have validated the method over the past 10 years for a variety of different oligonucleotides, such as RNA,^{52,53} DNA mismatches,⁵⁴ and DNA analogues such as TNA.⁵⁵ A recent study by our group,⁵⁶ using this type of mesoscopic modeling, found that the increased stability of LNA is largely hydrogen-bond-driven. Yet we also found that in certain stacking configurations, a decrease of stabilization may occur. These results were based on a small set of experimental melting temperatures, and we could not cover all possible stacking configurations. The sequences also had additional elements such as overhangs of variable length and, in some cases, fluorophores, which added to the uncertainty of our results. Therefore, it became necessary to recourse to a much larger set of melting temperatures, preferably without overhangs and fluorophores, which would enable us to identify the sources of LNA stabilization. Fortunately, the existing published data of single LNA modifications is unusually abundant,^{43,49,50} providing over 300 sequences and their melting temperatures and allowing us to cover all single LNA:DNA/DNA stacking interactions.

For probe design, one aims at maximizing the difference in probe hybridization temperatures between the mutated and wild-type genes. LNA enrichment allows us to further increase this difference,⁵⁷ but poorly placed LNA modifications may be of little help as we have shown recently.⁵⁶ The full mesoscopic

parameterization of LNA modifications opens the possibility to optimize the LNA inserts for oncogene probes. An improved stability and specificity of the probes is also highly desirable since it allows the detection of small quantities of the mutation, favoring an early diagnosis and detection. In terms of hybridization temperature prediction, the mesoscopic model has a similar predictive capability to the NN model;⁵⁸ however, it also allows stability analysis at a base-pair level, especially as it takes into account the nonlinear interactions between neighboring base pairs.

Here, we apply the new mesoscopic parameters to the high-throughput prediction of hybridization temperatures of the oncogene variants *BRAF*, *KRAS*, and *EGFR*, which were reported to be present in over 30% of all solid tumors, including breast, colon, and lung cancers.⁵⁹ Additionally, *KRAS* variants have been imposing a challenge in PCR and sequencing assays, resulting in smaller sensitivities.⁶⁰ It was suggested that the poor discrimination in *KRAS* probes is due to a CG internal region resulting in poor efficiency even in LNA-enriched probes.⁵⁶ First, we selected candidate probes for these three variants and then tested all possible combinations of one, two, and three LNA modifications, making use of the new mesoscopic parameters. From these, we selected six candidate probes, which were synthesized, had their melting temperatures measured, and we obtained an average prediction accuracy of 1 °C.

MATERIALS AND METHODS

Model. The Peyrard–Bishop (PB) mesoscopic model uses two different potentials in its Hamiltonian: a Morse potential representing the hydrogen bonds that connect each base and a harmonic potential describing the stacking interaction between adjacent base pairs

$$U_{i,i+1} = \frac{k_{\alpha,\gamma}}{2}(y_i - y_{i-1})^2 + D_\alpha(e^{-y/\lambda_\alpha} - 1)^2 \quad (1)$$

which describes the interaction of the i th base pair of type α with its nearest neighbor $i + 1$ of type γ . The parameters D_α and λ_α describe the depth and width of the potential well for the i th base pair of type α , respectively.

During our calculations, the potential width λ is kept constant; therefore, the potential D completely represents the Morse potential. For the remainder of this work, we use $\lambda = 3.2 \times 10^{-2}$ and 0.97×10^{-2} nm for base pairs adenine–thymine (AT) and cytosine–guanine (CG),⁵¹ respectively. Modified LNA base pairs use the same values for λ as their canonical analogues.

An elastic constant $k_{\alpha,\gamma}$ is used to describe the coupling between nearest-neighbor base pairs, and the coordinate y represents the relative displacements between the bases.

Equation 1 is used to calculate the partition function over all possible energetic configurations of N base pairs

$$Z_y = \int_{y_{\min}}^{y_{\max}} dy_1 \int_{y_{\min}}^{y_{\max}} dy_2 \dots \int_{y_{\min}}^{y_{\max}} dy_N \times \prod_{n=1}^N e^{-\beta U(y_n, y_{n+1})} \quad (2)$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, and T is the absolute temperature. The integral over all possible configurations of base-pair displacements, y_i , is performed and all possible Morse potentials, and stacking interactions are

considered simultaneously during the evaluation. The reduced degrees of freedom of the model represented by eq 2, while allowing for an efficient numerical evaluation, has a consequence of resulting in transition temperatures, which are very low for short sequences. To overcome this problem, we introduced an adimensional index τ , which is calculated from eq 2, and it is directly correlated with the experimental melting temperatures.⁶¹

Moreover, the average relative displacement, $\langle y_m \rangle$, at the m th position in the sequence can be derived from

$$\langle y_m \rangle = \frac{1}{Z_y} \int_{y_{\min}}^{y_{\max}} dy_1 \int_{y_{\min}}^{y_{\max}} dy_2 \dots \int_{y_{\min}}^{y_{\max}} dy_N \times y_m \prod_{n=1}^N e^{-\beta U(y_n, y_{n+1})} \quad (3)$$

Note that for the calculation of $\langle y_m \rangle$, we will use temperatures that are much lower than the actual melting temperatures.

Here, eqs 2 and 3 are evaluated via the transfer integral technique for heterogeneous sequences,⁶² and we use $y_{\min} = -0.1$ nm and $y_{\max} = 40$ nm, and a calculation temperature $T = 370$ K, which has no relation to the melting temperatures. The relative displacement y can be negative, meaning that the bases are moving toward each other. However, for negative y , the Morse potential is strongly repulsive, which is why a short $y_{\min} = -0.1$ nm is sufficient for numerical convergence. On the other hand, when the bases move away from each other (positive y), the Morse potential becomes flat and a large value of y is necessary to achieve numerical convergence.⁶² The index τ is largely temperature-independent, and $T = 370$ K was found to be adequate for its calculation from the classical partition function; for details, see ref 61.

Temperature Prediction. Considering a set

$$P = \{p_1, p_2, \dots, p_F\} \quad (4)$$

where p_i are Morse potentials, D_{ω} and stacking parameters, $k_{\alpha, \gamma}$. An index $\tau_i(P)$ is calculated for each sequence i using the partition function, eq 2, for the PB Hamiltonian. Thereon, the melting temperature, $T'_i(P)$, for each parameter set, P , is derived from the equation

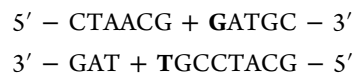
$$T'_i(P) = a_0(N) + a_1(N)\tau \quad (5)$$

where N is the length of the sequence. The calculation of the regression coefficients a_0 and a_1 is carried out at a total strand concentration C_t . The melting temperatures, $T'_i(P)$, required for the parameterization were obtained from published sources as described in the next section.

Notation. LNA modifications are preceded by a plus sign and marked in bold. Since antiparallel strands of native DNA duplexes exhibit structural symmetry, only two Morse potentials (AT, CG) are required to represent the hydrogen bonds between base pairs. Such symmetry does not apply to modified LNA base pairs, which leads to four new Morse parameters, A+T, T+A, C+G, and G+C, if we consider only base pairs of the type DNA+LNA:DNA or DNA:DNA+LNA. A similar situation occurs for the stacking parameters, some nearest-neighbor doublets are identical, for instance, ATpCG and GCpTA, such that 10 nearest-neighbor parameters are sufficient to represent internal DNA:DNA doublets. For example, the doublet GCpT+A is symmetrical to the +ATpCG doublet. The set of sequences collected in the literature contain all of the 32 possible single LNA

modifications, and the symmetry is reduced similarly for LNA NN and we always keep the NN that precede alphabetically. A full breakdown of the NN occurrences in the dataset is shown in Supporting Information Table S1.

As a practical example of the notation, consider the following sequence



in which case four parameters are required to represent the Morse potentials: AT, CG, A+T, and +GC, and 10 stacking potentials: CGpTA, TApAT, ATpA+T, A+TpCG, CGpGC, GCp+GC, +GCpAT, ATpTA, TApGC, and GCpCG.

Melting Temperature Set for Parameterization. The requirement for the sequences to be used for this work is that they contain only single LNA modification, that is, an LNA:DNA base pair. Further condition is that they should not have fluorophores attached and that any LNA:DNA nearest-neighbor configuration should be present in multiple sequences. Pure DNA:DNA control sequences, coming from the same LNA data sets were also used to adjust DNA parameters as we will describe in the next section. Existing data sets with tandem LNA modifications, such as Hughesman et al.,⁴² or with attached fluorophores, such as Owczarzy et al.,⁴¹ were not used for these reasons. In view of these requirements, we used the published melting temperatures for sequences measured at a high salt concentration (1021 mM Na⁺) collected from refs 43, 49, 50. A total of 306 sequences meet our requirements, ranging between 8 and 23 base pairs in length.

The regression scheme of the PB model, eq 5, requires that all melting temperatures should be at equivalent strand concentrations. Therefore, we used the strand concentration of 2 μ M, used in ref 50, as a reference. For the remaining sequences from refs 43, 49, we recalculated the melting temperatures using their respective reported total enthalpies and entropies. All sequences and their respective measured (or recalculated) and predicted melting temperatures are shown in Table S2.

Minimization Procedure. *Optimization.* To obtain an improved set of parameters for the PB model, we use an optimization method, detailed in refs 58, 63. Each parameter set P_j , eq 5, consisting of the parameters for the PB model will result in a melting temperature prediction $T'_i(P_j)$. This result is then compared with the experimental temperature T_i and the parameter set is varied until the squared difference is minimized

$$\chi_j^2 = \sum_{i=1}^N [T'_i(P_j) - T_i]^2 \quad (6)$$

Seed Parameters. This variation is initiated considering an initial set of i parameters, p_i . Those are varied in an interval

$$p_i \in [(1-f)s_i, (1+f)s_i] \quad (7)$$

that is, the value is sampled within a fraction $\pm f$ of a seed value for the parameter s_i . In this work, $f = 0.1$ results in the interval $[0.9s_i, 1.1s_i]$. The minimization procedure is numerically implemented using a downhill simplex method^{63,64} using eq 6 as the merit function, and its minimum is searched in the multidimensional space defined by the parameter set P_j . The

melting temperature deviation between the predicted and experimental temperatures is defined as

$$\langle \Delta T \rangle = \frac{1}{N} \sum_{i=1}^N |T'_i - T_i| \quad (8)$$

Initial Minimization (IM). In the first part of the minimization, IM, we kept constant all parameters without LNA modifications using previously published results for DNA at a high salt concentration, 1021 mM Na⁺,⁶³ that is, considering that DNA canonical parameters will remain approximately the same in the presence of LNA. Parameters with the modification are considered as having the same initial value as its nonmodified analogues. We let the minimization proceed for the 4 Morse and 32 stacking parameters that contain LNA modifications. A large number of minimization rounds did converge poorly; for this reason, we kept only 100 out of 300 rounds with the lowest χ^2 . Before this first optimization, the dataset had quality factors $\chi^2 = 4665.5$ °C² and $\langle \Delta T \rangle = 2.9$ °C. After optimization, they decreased to $\chi^2 = 1584.4$ °C², $\langle \Delta T \rangle = 1.7$ °C.

DNA/LNA Minimization (DL). In the previous steps, we have considered that the unmodified DNA bases keep the same parameters as from previous work.⁵¹ In this new minimization, we used IM as seed parameters, but now we also let the DNA base parameters vary and we will call this the DNA/LNA minimization (DL). That was followed for 300 minimization rounds, and the 100 with the lowest χ^2 were averaged as the final result.

Influence of the Experimental Error (EE-DL). To simulate the influence of experimental error (EE) associated with the temperature measurement on our new parameters, we change the temperature by small random amounts such that the standard deviation between the original set and the optimized set approaches the declared experimental uncertainty. Here, we used as the initial set of parameters the results from the minimization round DL, and therefore, this round is called EE-DL. As the whole dataset comes from three different sources, we used the highest declared uncertainty, 0.8° C, from Fakhfakh et al.⁴³ Again, the minimization procedure was followed 300 times, and we averaged the 100 with the lowest χ^2 as the final result. We use the standard deviation of the lowest 100 rounds to represent the uncertainty estimate of our new parameters. Each minimization step took approximately 7 days in a server with 400 cores, that is, the whole minimization took of the order of 21 days. A summary of the quality parameters for each minimization round is shown in Table 1.

Table 1. Summary of the Quality Parameters for the Minimization Rounds

round	$\langle \Delta T \rangle$ (°C)	χ^2 (°C ²)
IM	1.75	1584.4
DL	1.50	1053.7
EE-DL	1.47	1013.6

Capture/Linker Probe Design. Candidate DNA/DNA probes for *BRAF*, *EGFR*, and *KRAS* oncogenes were designed using public available DNA human genome sequencing data (NCBI) with GC content in the range of 38–55% and melting temperature above 45 °C according to the initial NN prediction for unmodified DNA in medium salt buffer. The latter criterion was applied to secure adequate binding

properties in most hybridization assays.^{65,66} Probe design was followed similarly as done before.^{67,68} Briefly, the assembly of the human oncogenes and their respective NCBI code are described in Table S4.

Mutated oncogenes were assessed using dbSNP base (rs113488022) (*BRAF* V600E); rs121913529 (*KRAS* G12D); rs112445441 (*KRAS* G13D); rs121434568 (*EGFR* L858R). Oligonucleotide probes were designed to be complementary to the position of the corresponding gene bearing the mutation (capture probe), and downstream the gene with a gap of >20 nt from the position of capture probe, for linker probe. The length of each probe was determined using probe uniqueness software described in ref 69.

Once the candidate capture (cap) and linker (lin) probes were established, we used the optimized parameters DNA + LNA:DNA parameters to calculated all possible configurations with one, two, or three LNA modifications for each of the selected probes. The melting temperatures of both, modified and nonmodified probes, were evaluated, and a salt correction from Owczarzy et al.⁷⁰ to a salt concentration of 137 mM Na⁺ was applied to the results. We calculated between 834 and 2345 different LNA-modified sequences, depending on length, for each candidate probe. Supporting Information Tables S5–S11 show the 30 highest and 30 lowest melting temperature predictions for *BRAF*, *EGFR*, and *KRAS* linker and capture variants, respectively.

We selected the probes with the lowest possible number of LNA incorporation per sequence to achieve the highest overall T_m and the highest discrimination T_m full match vs T_m mismatch for the probes binding to the position of mutation in the corresponding oncogene. Selected candidate probes were synthesized and their melting temperatures were measured as described in the next section.

Oligonucleotide Synthesis and UV Melting Procedure. The LNA/DNA oligonucleotides were synthesized on an ASM-800 ET synthesizer from BIOSSET Ltd. Commercial phosphoramidites from Sigma-Aldrich (dA(Bz), dC(Bz), dG(ib), dT) and Qiagen (LNA-A(Bz), LNA-T, LNA-mC(Bz), LNA-G(dmf)) were dissolved in dry acetonitrile to a concentration of 0.07 M. Reagents for solid-phase synthesis were purchased from Sigma-Aldrich: TCA Deblock, DCI activator 0.25 M, Oxidizer 0.02 M, Cap A, and Cap B. The phosphoramidites were loaded onto the synthesizer following the manufacture standard protocol. All of the oligonucleotides were synthesized on controlled pore glass (CPG 1000 Å) universal support purchased from Sigma-Aldrich in a 1 μmol scale using a double-coupling protocol. DMT-off mode was applied. The coupling rate during the synthesis was estimated based on absorbance measurements of the DMT cation using TM800 software.

After synthesis, the columns were dried by nitrogen purge and then transferred to 5 mL tubes. For cleavage from the universal support, aqueous ammonia (28–30%; 1 mL) was added and the samples were placed at 55 °C for 12 h. Next, the samples were cooled at –20 °C for about 10 min and filtrated into 2 mL Eppendorf tubes. The oligonucleotides were evaporated to dryness and redissolved in 1000 μL of Milli-Q water.

The identity of oligonucleotides was confirmed by mass spectrometry (MS) using an Autoflex speed MALDI-TOF mass spectrometer (Bruker Daltonics, Germany). Matrix-assisted laser desorption ionization–mass spectrometry (MALDI-MS) of purified oligonucleotides and their respective

purities are shown in Supporting Information Table S12. Representative MALDI-MS spectra for each probe are shown in Supporting Information Figures S2–S14. Samples were co-spotted with 3-hydroxypicolinic acid as matrix on an MTP AnchorChip target plate for the analysis. The obtained mass spectra were recorded by the flexControl 3.4 (Bruker Daltonics, Germany) software. The oligonucleotides were purified on an Ultimate 3000 UHPLC (Dionex, United States) using a DNA-Pac RP (Thermo Fisher Scientific, United States) column (4 μm , 3.0 \times 100 mm) with a gradient of 5–25% buffer B in A over 30 min at 60 $^{\circ}\text{C}$ (buffer A: 0.05 M TEAA, buffer B: 25% A in acetonitrile). The peaks were monitored at 260 nm. Representative ultrahigh-performance liquid chromatography (UHPLC) traces for each probe are shown in Supporting Information Figures S1–S13.

The melting temperature studies were performed on a DU800 UV/VIS spectrophotometer equipped with a Beckman Coulter Performance Temperature Controller. Complementary strands (0.5 μM of each strand), in 1 \times phosphate-buffered saline (PBS) were mixed, denatured for 10 min at 90 $^{\circ}\text{C}$, and subsequently cooled to 15 $^{\circ}\text{C}$, the temperature at which the experiment was started. Reported melting temperatures present the maximum of the first derivative of the curve and are an average of the two measurements within a deviation of 1 $^{\circ}\text{C}$. T_{m} curves for each probe are shown in Supporting Information Figures S15–S20.

RESULTS AND DISCUSSION

All results presented here refer to the final round, EE-DL, unless noted otherwise.

Morse Potentials. The Morse potential has a long history for the use of modeling hydrogen bonding in molecular dynamics,⁷¹ coarse-grained models,⁷² and for mesoscopic models.⁷³ Other model potentials, such as the Lennard-Jones potential, could, in principle, be used, yet no appreciable difference was noted compared to the Morse potential in the PB model.⁷⁴ For the remainder of this discussion, we will refer to the potential depth D simply as the Morse potential.

In Figure 1, we show the final average parameter D of the Morse potentials and, for comparison, the previous results for canonical DNA.⁵¹ Modified base pairs consistently show higher Morse potentials than their analogues. This suggests that, in comparison to the unmodified DNA, new hydrogen

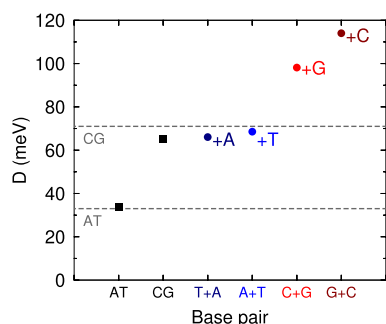


Figure 1. Average D parameter of the Morse potential for LNA-modified (bullets) and unmodified DNA (boxes) base pairs. LNA modifications are shown next to symbols. Error bars were omitted as they are smaller than the symbol size. For comparison, we show the published parameters⁵¹ for unmodified DNA base pairs, AT and CG, as gray dashed lines. For clarity, LNA monomers are separated by color, and will use this color scheme in the next figure.

bonds between the base pairs may have been formed due to the LNA modification, or that the existing ones were strengthened in some way. Morse potentials of unmodified DNA base pairs, AT and CG, remained nearly the same as that of previous calculations,^{51,75} shown as dashed lines in Figure 1, which is consistent with the similarity of buffer conditions between all data sets. As we are dealing with single LNA modification in a DNA duplex, little influence of LNA on the remaining base pairs was expected. The small change of the canonical base-pair parameters also confirms this and is consistent with atomistic molecular dynamics simulations, which found that the change induced by single LNA modification in a DNA oligonucleotide is very localized and essentially limited to the immediate neighboring base pairs.⁷⁶ It was suggested by a thermodynamic study a similar stability ordering as the one we have found here for Morse potentials, $+A \leq +T \leq +G \leq +C$.⁴⁹ Although we have found a stronger Morse potential for the modified guanine, $+G$. However, we cannot establish a direct comparison since the stability on the PB model is also related to the stacking parameters. One of the consequences of the stronger Morse potentials is a reduction in entropy, which can be calculated from the partition function eq 2. This is consistent with the entropy reduction seen for nearest-neighbor analysis.^{42,43} Some examples of calculated configurational entropies with LNA-modified sequences are shown in Supporting Information Figure S25.

There is an ongoing debate regarding the intramolecular origin of the increased stability of LNA modifications. On the one hand, there have been proposals that LNA stability is due to increased stacking interactions.^{44,45,49,77} On the other hand, Jensen et al.⁴⁸ concluded against either increased stacking or hydrogen bonding and argued that hybridized water may be the cause of LNA stability. An NMR study by Egli et al.⁴⁶ established that the extra oxygen atoms in LNA monomers are engaged in water-bridged hydrogen bonds. This may account for an increased thermodynamic stability as an extensive hydration of hydrogen-bond acceptors and donors is associated with a stability increase.⁴⁶ In another molecular dynamics study,⁷⁸ modified LNA duplexes were found forming even four water-bridged hydrogen bonds. In this study,⁷⁸ LNA helices were found to be less hydrated than DNA or RNA, which was hypothesized to be due to its shorter intrastrand phosphate distances and lower backbone flexibility, allowing the formation of water-mediated hydrogen bonds providing extra stability to LNA base pairs.

Our results undoubtedly support the idea of increased hydrogen bonding, as shown in Figure 1, which is also in line with our earlier findings⁵⁶ for a much smaller dataset. It is particularly interesting to note that both AT modifications, A +T and T+A, have become almost identical in strength to an unmodified CG base pair (Figure 1). There are suggestions that AT may have a third weak bond $\text{C}-\text{H}\cdots\text{O}$,^{79–81} although it may not be a hydrogen bond but a van der Waals interaction.⁸² While our calculations cannot pinpoint the exact nature of this interaction, one might speculate that the $\text{C}-\text{H}\cdots\text{O}$ contact may have been strengthened due to the LNA modification, for both +A and +T.

Stacking Potentials. The calculated stacking parameters are shown in Figure 2. Higher fluctuations from the canonical values depend on the modified base-pair location and direction on the strand. Similarly to previous results for DNA,⁵¹ NNs containing only AT base pairs (Figure 2a) present larger stacking variations. Less fluctuation is seen for mixed NNs

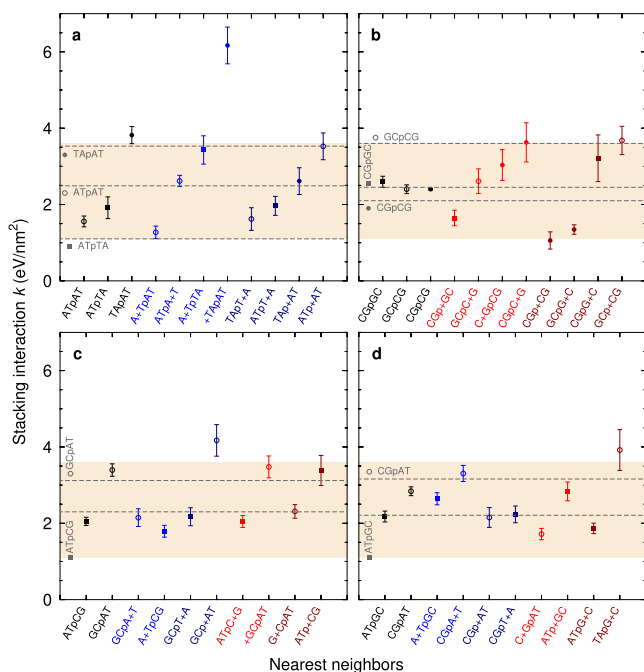


Figure 2. Average stacking potentials for the LNA-modified and canonical nearest neighbors. Results are separated into panels grouped by NN patterns: (a) for ATpTA (boxes), ATpAT (circles), and TApAT (bullets); (b) for CGpCG (bullets), CGpCG (boxes), and GCpCG (circles); (c) for ATpCG (boxes) and GCpAT (circles); (d) for ATpGC (boxes) and CGpAT (circles). Error bars were estimated in the EE-DL minimization round. For comparison, we show the published stacking potentials for canonical DNA⁵¹ as gray dashed lines. The shaded region shows the range of all possible stacking parameters for canonical DNA,⁵¹ with the minimum being the stacking parameter for ATpTA and maximum for GCpCG (the same on all panels). LNA monomers are separated by color and follow the same color scheme as Figure 1.

(Figure 2c,d), which tend to fluctuate around the stacking parameters of their unmodified analogues. Molecular dynamics studies have established a local disturbance due to LNA base pairs affecting the first nearest neighbors,⁷⁶ which is similarly related to the little change in canonical interactions found in our results. Moreover, LNA enrichment does not have a strong effect on canonical stacking interactions, except for CG shown in Figure 2b, as they keep nearly the same within the standard deviation during the minimization. Only one LNA-modified NN, +TApAT, displayed significantly larger stacking, com-

pletely out of the region of canonical stacking (shaded area in Figure 2a).

Figure 2 shows that very few LNA modifications have stacking interactions that are larger than typical DNA stacking, that is, that lie above the shaded area. Only one type of modification, +TApAT, has a very strong stacking interaction, which may indicate a specially modified stacking arrangement⁵⁴ or perhaps to a departure from the *anti-anti* configuration.⁸³ In fact, a considerable number of LNA:DNA configurations decrease their stacking interactions in comparison to their unmodified analogues. Therefore, we conclude that, in general, stacking does not account for the increased stability of LNA, and in many cases, it appears to be even a destabilizing factor. Incidentally, the single sequence analyzed by Nielsen et al.,⁴⁴ who concluded stacking as a major contributing factor to LNA stability, contains +TApAT, which is the only strongly stacked modification that we observed (Figure 2). Therefore, even though we conclude against stacking as a major factor of LNA:DNA stability, our results are not at odds with the conclusion of Nielsen et al.⁴⁴ for +TApAT.

Probe Design, Synthesis, and Evaluation. The calculation of the LNA:DNA parameters is time-consuming and requires considerable computational resources. However, once they become available, it is straightforward to calculate melting temperature predictions for a large set of sequences. For prediction, the computational efficiency of the PB regression scheme is comparable to the NN model.^{84,85} First, candidate probes were selected from genomic analysis for *BRAF*, *EGFR*, and *KRAS* oncogenes. Then, with the new LNA:DNA parameters at hand, we calculated the melting temperatures of LNA modifications at all positions of *BRAF*, *EGFR*, and *KRAS* capture and linker probes. The resulting temperatures, which are valid for high sodium concentrations, were recalculated by applying a salt correction factor from Owczarzy et al.,⁷⁰ which was shown to be adequate for LNA.⁵⁰ An excerpt of the predicted melting temperatures for up to three candidate LNA modifications per probe is shown in Supporting Information Tables S5–S11.

One probe was selected from the list of the candidate probes for each capture and linker pair, using additional criteria of the lowest possible number of LNA incorporations per sequence to achieve the highest overall T_m and the highest discrimination T_m full match vs T_m mismatch for the probes binding to the position of mutation in the corresponding oncogene. The selected probes were synthesized, and their

Table 2. Measured T_{exp} , Control T_{ctrl} , and Predicted Temperatures T_{pred} for the Selected Sequences from the Probe Prediction^a

ID	synthesized probe with LNA modifications	T_{ctrl}	T_{exp}	T_{pred}
BRAFcap-M1225	ATCGAGAT+TTCT+CTGTAG+CTA	59.57	64.50	64.92
BRAFlin-M416	CAA+CTGTT+CAAA+CTGAT	50.75	62.20	60.75
KRAS12cap-M1344	GCACTCTTGCCTACCCA+ATC	64.07	64.70	65.42
KRAS13cap-M1304	GCACTCTTGCCTA+CGCA+TTC	64.15	68.90	68.83
KRASlin-M616	TGAAGT+CA+CA+TTATATA	48.15	55.00	58.17
EGFRcap-M943	GAG+AAAAAGTTTCTCA+TGTA+CAGT	57.40	61.2	61.76
EGFRlin-M478	TTG+TTGGAT+CATA+TTCGT	54.45	61.50	61.52

$$\langle T_{exp} - T_{pred} \rangle = 0.91 \text{ } ^\circ\text{C}$$

^aControl T_{ctrl} refers to the unmodified DNA/DNA probes. The probe identification (ID) refers to the candidate probe list in Supporting Information Tables S5–S11. The average deviation $\langle T_{exp} - T_{pred} \rangle$ was calculated between the available predicted and measured temperatures. Probe sequences are shown in the 5' → 3' direction.

melting temperatures were measured as described in the [Materials and Methods](#) section. [Table 2](#) shows the predicted and measured melting temperatures for each probe, and their average deviation was found to be of the order of 1 °C. While this represents an overall satisfactory prediction, it is worthwhile to have a closer look at the probes that fell short of their prediction target. For this purpose, we calculated the opening profiles of these sequences using [eq 3](#), which provides an intuitive way to visualize localized instabilities in the sequence. For instance, KRASlin-M616 was predicted to melt at 3 °C higher than measured. [Figure 3](#) shows how this probe

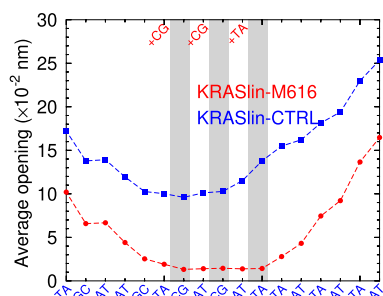


Figure 3. Average opening profile for the LNA-modified KRASlin-M616 probe (red bullets) and its corresponding unmodified control probe (blue squares). The location of the LNAs is indicated by the gray shaded area. Calculation was carried out at 220 K, [eq 3](#), which has no relation to the melting temperature.

stabilizes (lower openings) in comparison to its control sequence (larger openings). However, the added LNA modifications provided only moderate additional stabilization for its 3' AT-tail. One possibility for the lower measured melting temperature is that the high salt parameters overestimate the stability of AT base pairs at 3' sequence ends. In previous studies,⁷⁵ we observed that AT base pairs at the sequence end tended to have much lower Morse potentials for low salt concentrations; it is therefore possible that the lower-than-predicted temperature has a similar effect. For BRAFlin-M416, the measured temperature was 1 °C higher than the predicted value. While this is within the measurement uncertainty, [Figure 4](#) shows that the distribution of +C along

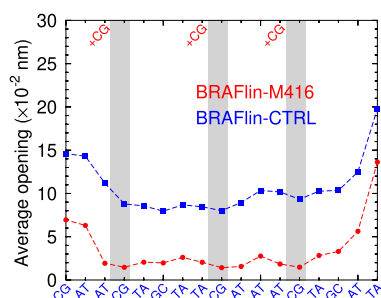


Figure 4. Average opening profile for the LNA-modified BRAFlin-M416 probe (red bullets) and its corresponding unmodified control probe (blue squares). Figure elements are the same as for [Figure 3](#).

the sequence has a very large stabilizing effect. KRAS12cap-M1344 was selected among the lowest-stabilizing LNA modifications, and indeed its measured melting temperature is only marginally larger than that of the control sequence. [Figure 5](#) shows a small stabilization end toward the 3' end,

which is already capped by a CG base pair, and comparing the +AT openings, one notices its similarity to an unmodified CG.

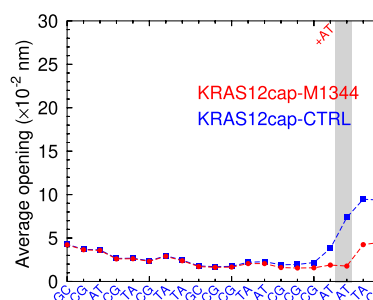


Figure 5. Average opening profile for the LNA-modified KRAS12cap-M1344 (red bullets) and its corresponding unmodified control probe (blue squares). Figure elements are the same as for [Figure 3](#).

CONCLUSIONS

We have successfully applied a mesoscopic model to parameterize a large set of published melting temperatures for LNA-modified sequences measured at a high salt concentration. We have found a substantial increase in Morse potentials, indicating stronger hydrogen bonding for LNA:DNA base pairs. This confirms previous results at low salt concentrations⁵⁶ and establishes hydrogen bonding as the main source of LNA:DNA stabilization. The large number of sequences allowed us to fully parameterize all 32 possible LNA:DNA/DNA nearest-neighbor stacking interaction. Only one case of increased stacking for +TApAT was found, while the remaining ones are similar to unmodified DNA with a few cases of reduced stacking interactions. The new parameters were used to predict probes hybridization, targeting the oncogene variants *BRAF*, *KRAS*, and *EGFR* in a medium salt buffer, as a proof of concept. For each probe, we tested all combinations of up to three LNA modifications and selected six candidate probes for synthesis and melting temperature measurements. We have found a good agreement in the predictions, after applying established salt correction factors.⁷⁰ The few discrepancies between prediction and measured temperatures were analyzed using opening profiles at a base-pair level.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00470>.

Number of modified and nonmodified NN ([Table S1](#)); sequences used in the minimization and their respective measured and predicted temperatures ([Table S2](#)); measured and predicted melting temperatures for the validation set of sequences containing LNA tandem modifications ([Table S3](#)); summary of the oncogenes used for the probe design and their respective human assemblies ([Table S4](#)); melting temperature predictions for *BRAF*, *EGFR*, and *KRAS* variants ([Tables S5–S11](#)); MALDI-MS and purities of the measured probes ([Tables S12](#)); HPLC and MALDI-MS of the measured probes ([Figures S1–S14](#)); representative T_m curve of the measured probes ([Figures S15–S20](#)); average opening profiles for the LNA-modified probes measured in this study ([Figures S21–S24](#)); and some examples of

calculated entropy for LNA-modified sequences (Figure S25) (PDF)

AUTHOR INFORMATION

Corresponding Authors

Kira Astakhova – Department of Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark; orcid.org/0000-0003-4878-0301; Email: kiraas@kemi.dtu.dk

Gerald Weber – Departamento de Física, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil; orcid.org/0000-0002-2935-1571; Email: gweber@ufmg.br

Authors

Izabela Ferreira – Departamento de Física, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil; Programa Interunidades de Pós-Graduação em Bioinformática, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil

Sofie Slott – Department of Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jcim.1c00470>

Notes

The authors declare no competing financial interest. All sequences, their respective melting temperatures, and all calculated parameters are included in our updated software package TfrReg (<https://bioinf.fisica.ufmg.br/software>) and available as buildable packages in the OpenSuse repository (<https://build.opensuse.org/package/show/home:drgrweber/TfrReg>) and are free of charge. All results presented here can be verified with this software. This package also calculates opening profiles, such as those of Figures 3–5, which take only a few seconds to run on standard computers with the provided parameters.

ACKNOWLEDGMENTS

I.F. and G.W. were supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes, Brazil, Finance Code 001). K.A. and S.S. were supported by DTU Ellitte Ph.D. Scholarship and DTU Proof of concept programme.

REFERENCES

- (1) Obika, S.; Nanbu, D.; Hari, Y.; Morio, K.-i.; In, Y.; Ishida, T.; Imanishi, T. Synthesis of 2'-O, 4'-C-methylenuridine and-cytidine. Novel bicyclic nucleosides having a fixed C3-endo sugar pucker. *Tetrahedron Lett.* **1997**, *38*, 8735–8738.
- (2) Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Olsen, C. E.; Wengel, J. LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron* **1998**, *54*, 3607–3630.
- (3) Kumar, R.; Singh, S. K.; Koshkin, A. A.; Rajwanshi, V. K.; Meldgaard, M.; Wengel, J. The first analogues of LNA (locked nucleic acids): phosphorothioate-LNA and 2'-thio-LNA. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2219–2222.
- (4) Obika, S.; Nanbu, D.; Hari, Y.; Andoh, J.-i.; Morio, K.-i.; Doi, T.; Imanishi, T. Stability and structural features of the duplexes containing nucleoside analogues with a fixed N-type conformation,

2'-O, 4'-C-methylenuridine nucleosides. *Tetrahedron Lett.* **1998**, *39*, 5401–5404.

(5) Singh, S. K.; Koshkin, A. A.; Wengel, J.; Nielsen, P. LNA (Locked nucleic acids) synthesis and high-affinity nucleic acid recognition. *Chem. Commun.* **1998**, *4*, 455–456.

(6) Wang, G.; Gunic, E.; Girardet, J.-L.; Stoisavljevic, V. Conformationally locked nucleosides. Synthesis and hybridization properties of oligodeoxynucleotides containing 2',4'-C-bridged 2'-deoxynucleosides. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1147–1150.

(7) Obika, S.; Hari, Y.; Morio, K.-i.; Imanishi, T. Triplex formation by an oligonucleotide containing conformationally locked C-nucleoside, 5-(2-O, 4-C-methylene- β -D-ribofuranosyl) oxazole. *Tetrahedron Lett.* **2000**, *41*, 221–224.

(8) Wahlestedt, C.; Salmi, P.; Good, L.; Kela, J.; Johnsson, T.; Hökfelt, T.; Broberger, C.; Porreca, F.; Lai, J.; Ren, K. Potent and nontoxic oligonucleotides containing locked nucleic acids. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5633–5638.

(9) Pasternak, A.; Wengel, J. Thermodynamics of RNA duplexes modified with unlocked nucleic acid nucleotides. *Nucleic Acids Res.* **2010**, *38*, 6697–6706.

(10) Kaur, H.; Arora, A.; Wengel, J.; Maiti, S. Thermodynamic, counterion, and hydration effects for the incorporation of locked nucleic acid nucleotides into DNA duplexes. *Biochemistry* **2006**, *45*, 7347–7355.

(11) Kurreck, J.; Wyszko, E.; Gillen, C.; Erdmann, V. A. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res.* **2002**, *30*, 1911–1918.

(12) Jacobsen, N.; Bentzen, J.; Meldgaard, M.; Jakobsen, M. H.; Fenger, M.; Kauppinen, S.; Skouv, J. LNA-enhanced detection of single nucleotide polymorphisms in the apolipoprotein E. *Nucleic Acids Res.* **2002**, *30*, No. e100.

(13) Josefson, M. H.; Löfström, C.; Sommer, H. M.; Hoorfar, J. Diagnostic PCR: comparative sensitivity of four probe chemistries. *Mol. Cell. Probes* **2009**, *23*, 201–203.

(14) Morandi, L.; De Biase, D.; Visani, M.; Cesari, V.; De Maglio, G.; Pizzolitto, S.; Pession, A.; Tallini, G. Allele specific locked nucleic acid quantitative PCR (ASLNAqPCR): an accurate and cost-effective assay to diagnose and quantify KRAS and BRAF mutation. *PLoS One* **2012**, *7*, No. e36084.

(15) Fontanilles, M.; Marguet, F.; Ruminy, P.; Basset, C.; Noel, A.; Beaussire, L.; Viennot, M.; Vially, P.-J.; Cassinari, K.; Chambon, P. Simultaneous detection of EGFR amplification and EGFRvIII variant using digital PCR-based method in glioblastoma. *Acta Neuropathol. Commun.* **2020**, *8*, No. 52.

(16) Astakhova, I. V.; Ustinov, A. V.; Korshun, V. A.; Wengel, J. LNA for optimization of fluorescent oligonucleotide probes: improved spectral properties and target binding. *Bioconjugate Chem.* **2011**, *22*, 533–539.

(17) Østergaard, M. E.; Cheguru, P.; Papasani, M. R.; Hill, R. A.; Hrdlicka, P. J. Glowing locked nucleic acids: brightly fluorescent probes for detection of nucleic acids in cells. *J. Am. Chem. Soc.* **2010**, *132*, 14221–14228.

(18) Hussung, S.; Follo, M.; Klar, R. F.; Michalczyk, S.; Fritsch, K.; Nollmann, F.; Hipp, J.; Duyster, J.; Scherer, F.; von Bubnoff, N. Development and clinical validation of discriminatory multi-target digital droplet PCR assays for the detection of hot spot KRAS and NRAS mutations in cell-free DNA. *J. Mol. Diagn.* **2020**, *22*, 943–956.

(19) Bakthavathsalam, P.; Longatte, G.; Jensen, S. O.; Manefield, M.; Gooding, J. J. Locked nucleic acid molecular beacon for multiplex detection of loop mediated isothermal amplification. *Sens. Actuators, B* **2018**, *268*, 255–263.

(20) Bagheri, P.; Sharifi, M.; Ghadiri, A. Downregulation of MIR100HG Induces Apoptosis in Human Megakaryoblastic Leukemia Cells. *Indian J. Hematol. Blood Transfus.* **2021**, *37*, 232–239.

(21) Zhang, K.; Zheludev, I. N.; Hagey, R. J.; Wu, M. T.-P.; Haslecker, R.; Hou, Y. J.; Kretsch, R.; Pintilie, G. D.; Rangan, R.; Kladwang, W. Cryo-electron Microscopy and Exploratory Antisense

Targeting of the 28-kDa Frameshift Stimulation Element from the SARS-CoV-2 RNA Genome. *bioRxiv* **2020**, 395, No. 1245.

- (22) Lim, K. R. Q.; Maruyama, R.; Echigoya, Y.; Nguyen, Q.; Zhang, A.; Khawaja, H.; Sen Chandra, S.; Jones, T.; Jones, P.; Chen, Y.-W.; Yokota, T. Inhibition of DUX4 expression with antisense LNA gapmers as a therapy for facioscapulohumeral muscular dystrophy. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 16509–16515.
- (23) Schmidt, K. S.; Borkowski, S.; Kurreck, J.; Stephens, A. W.; Bald, R.; Hecht, M.; Friebe, M.; Dinkelborg, L.; Erdmann, V. A. Application of locked nucleic acids to improve aptamer in vivo stability and targeting function. *Nucleic Acids Res.* **2004**, *32*, 5757–5765.
- (24) Darfeuille, F.; Reigadas, S.; Hansen, J. B.; Orum, H.; Di Primo, C.; Toulmé, J.-J. Aptamers targeted to an RNA hairpin show improved specificity compared to that of complementary oligonucleotides. *Biochemistry* **2006**, *45*, 12076–12082.
- (25) Wojtyniak, M.; Schmidtgal, B.; Kirsch, P.; Ducho, C. Towards Zwitterionic Oligonucleotides with Improved Properties: the NAA/LNA-Gapmer Approach. *ChemBioChem* **2020**, *21*, 3234–3243.
- (26) Braasch, D. A.; Jensen, S.; Liu, Y.; Kaur, K.; Arar, K.; White, M. A.; Corey, D. R. RNA interference in mammalian cells by chemically-modified RNA. *Biochemistry* **2003**, *42*, 7967–7975.
- (27) Elmén, J.; Thonberg, H.; Ljungberg, K.; Frieden, M.; Westergaard, M.; Xu, Y.; Wahren, B.; Liang, Z.; Ørum, H.; Koch, T. Locked nucleic acid (LNA) mediated improvements in siRNA stability and functionality. *Nucleic Acids Res.* **2005**, *33*, 439–447.
- (28) Vester, B.; Lundberg, L. B.; Sørensen, M. D.; Babu, B. R.; Douthwaite, S.; Wengel, J. LNAs: Incorporation of LNA-Type Monomers into DNAs Markedly Increases RNA Cleavage. *J. Am. Chem. Soc.* **2002**, *124*, 13682–13683.
- (29) Jakobsen, M. R.; Haasnoot, J.; Wengel, J.; Berkhout, B.; Kjems, J. Efficient inhibition of HIV-1 expression by LNA modified antisense oligonucleotides and DNAs targeted to functionally selected binding sites. *Retrovirology* **2007**, *4*, No. 29.
- (30) Jadhav, V. M.; Scaria, V.; Maiti, S. Antagomirzymes: oligonucleotide enzymes that specifically silence microRNA function. *Angew. Chem., Int. Ed.* **2009**, *48*, 2557–2560.
- (31) Wang, L.; Yang, C. J.; Medley, C. D.; Benner, S. A.; Tan, W. Locked nucleic acid molecular beacons. *J. Am. Chem. Soc.* **2005**, *127*, 15664–15665.
- (32) Wienholds, E.; Kloosterman, W. P.; Miska, E.; Alvarez-Saavedra, E.; Berezikov, E.; de Bruijn, E.; Horvitz, H. R.; Kauppinen, S.; Plasterk, R. H. MicroRNA expression in zebrafish embryonic development. *Science* **2005**, *309*, 310–311.
- (33) Thomsen, R.; Nielsen, P. S.; Jensen, T. H. Dramatically improved RNA in situ hybridization signals using LNA-modified probes. *RNA* **2005**, *11*, 1745–1748.
- (34) Elmén, J.; Lindow, M.; Schütz, S.; Lawrence, M.; Petri, A.; Obad, S.; Lindholm, M.; Hedtjärn, M.; Hansen, H. F.; Berger, U. LNA-mediated microRNA silencing in non-human primates. *Nature* **2008**, *452*, 896–899.
- (35) Huang, S.; Ichikawa, Y.; Yoshitake, K.; Igarashi, Y.; Kinoshita, S.; Asaduzzaman, M.; Omori, F.; Maeyama, K.; Nagai, K.; Watabe, S. Potential silencing of gene expression by PIWI-interacting RNAs (piRNAs) in somatic tissues in mollusk. *bioRxiv* **2020**, No. 199877.
- (36) Kalinina, M.; Skvortsov, D.; Kalmykova, S.; Ivanov, T.; Dontsova, O.; Pervouchine, D. D. Multiple competing RNA structures dynamically control alternative splicing in the human ATE1 gene. *Nucleic Acids Res.* **2021**, *49*, 479–490.
- (37) Ignatiadis, M.; Sledge, G. W.; Jeffrey, S. S. Liquid biopsy enters the clinic—implementation issues and future challenges. *Nat. Rev. Clin. Oncol.* **2021**, 297–312.
- (38) Van Der Pol, Y.; Mouliere, F. Toward the early detection of cancer by decoding the epigenetic and environmental fingerprints of cell-free DNA. *Cancer Cell* **2019**, *36*, 350–368.
- (39) Liu, X.-P.; Yin, X.-H.; Meng, X.-Y.; Yan, X.-H.; Wang, F.; He, L. Development and validation of a 9-gene prognostic signature in patients with multiple myeloma. *Front. Oncol.* **2019**, *8*, No. 615.
- (40) Bruylants, G.; Bocconcelli, M.; Snoussi, K.; Bartik, K. Comparison of the thermodynamics and base-pair dynamics of a full LNA:DNA duplex and of the isosequential DNA:DNA duplex. *Biochemistry* **2009**, *48*, 8473–8482.
- (41) Owczarzy, R.; You, Y.; Groth, C.; Tataurov, A. Stability and Mismatch Discrimination of Locked Nucleic Acid-DNA Duplexes. *Biochemistry* **2011**, *50*, 9352–9367.
- (42) Hughesman, C. B.; Turner, R. F.; Haynes, C. A. Role of the heat capacity change in understanding and modeling melting thermodynamics of complementary duplexes containing standard and nucleobase-modified LNA. *Biochemistry* **2011**, *50*, 5354–5368.
- (43) Fakhfakh, K.; Marais, O.; Cheng, X. B. J.; Castañeda, J. R.; Hughesman, C. B.; Haynes, C. Molecular thermodynamics of LNA:DNA base pairs and the hyperstabilizing effect of 5'-proximal LNA:DNA base pairs. *AIChE J.* **2015**, *61*, 2711–2731.
- (44) Nielsen, C. B.; Singh, S. K.; Wengel, J.; Jacobsen, J. P. The solution structure of a locked nucleic acid (LNA) hybridized to DNA. *J. Biomol. Struct. Dyn.* **1999**, *17*, 175–191.
- (45) Nielsen, K. E.; Singh, S. K.; Wengel, J.; Jacobsen, J. P. Solution structure of an LNA hybridized to DNA: NMR study of the d(CTLGCTLTCTLCTGC):d(GCAGAAGCAG) duplex containing four locked nucleotides. *Bioconjugate Chem.* **2000**, *11*, 228–238.
- (46) Egli, M.; Minasov, G.; Teplova, M.; Kumar, R.; Wengel, J. X-ray crystal structure of a locked nucleic acid (LNA) duplex composed of a palindromic 10-mer DNA strand containing one LNA thymine monomer. *Chem. Commun.* **2001**, 651–652.
- (47) Eichert, A.; Behling, K.; Betzel, C.; Erdmann, V. A.; Fürste, J. P.; Förster, C. The crystal structure of an "All Locked" nucleic acid duplex. *Nucleic Acids Res.* **2010**, *38*, 6729–6736.
- (48) Jensen, G. A.; Singh, S. K.; Kumar, R.; Wengel, J.; Jacobsen, J. P. A comparison of the solution structures of an LNA:DNA duplex and the unmodified DNA:DNA duplex. *J. Chem. Soc., Perkin Trans. 2* **2001**, *2*, 1224–1232.
- (49) McTigue, P. M.; Peterson, R. J.; Kahn, J. D. Sequence-dependent thermodynamic parameters for Locked Nucleic Acid (LNA)-DNA duplex formation. *Biochemistry* **2004**, *43*, 5388–5405.
- (50) You, Y.; Moreira, B. G.; Behlke, M. A.; Owczarzy, R. Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Res.* **2006**, *34*, No. e60.
- (51) Weber, G.; Essex, J. W.; Neylon, C. Probing the Microscopic Flexibility of DNA from Melting Temperatures. *Nat. Phys.* **2009**, *5*, 769–773.
- (52) Amarante, T. D.; Weber, G. Evaluating hydrogen bonds and base stackings of single, tandem and terminal GU in RNA mismatches with a mesoscopic model. *J. Chem. Inf. Model.* **2016**, *56*, 101–109.
- (53) Ferreira, I.; Amarante, T. D.; Weber, G. Salt dependent mesoscopic model for RNA with multiple strand concentrations. *Biophys. Chem.* **2020**, 271, No. 106551.
- (54) Oliveira, L. M.; Long, A. S.; Brown, T.; Fox, K. R.; Weber, G. Melting temperature measurement and mesoscopic evaluation of single, double and triple DNA Mismatches. *Chem. Sci.* **2020**, *11*, 8273–8287.
- (55) Muniz, M. I.; Lackey, H. H.; Heemstra, J. M.; Weber, G. DNA/TNA mesoscopic modeling of melting temperatures suggest weaker hydrogen bonding of CG than in DNA/RNA. *Chem. Phys. Lett.* **2020**, No. 137413.
- (56) Domljanovic, I.; Taskova, M.; Miranda, P.; Weber, G.; Astakhova, K. Nucleic acid probes—Optical and theoretical study reveals new details on strand Recognition. *Commun. Chem.* **2020**, *3*, No. 111.
- (57) Tolstrup, N.; Nielsen, P. S.; Kolberg, J. G.; Frankel, A. M.; Vissing, H.; Kauppinen, S. OligoDesign: optimal design of LNA (locked nucleic acid) oligonucleotide capture probes for gene expression profiling. *Nucleic Acids Res.* **2003**, *31*, 3758–3762.
- (58) Weber, G. TifReg: Calculating DNA and RNA Melting Temperatures and Opening Profiles with Mesoscopic Models. *Bioinformatics* **2013**, *29*, 1345–1347.
- (59) Cheng, M. L.; Pectasides, E.; Hanna, G. J.; Parsons, H. A.; Choudhury, A. D.; Oxnard, G. R. Circulating tumor DNA in advanced

solid tumors: Clinical relevance and future directions. *Ca-Cancer J. Clin.* **2020**, *71*, 176–190.

(60) Yamada, T.; Iwai, T.; Takahashi, G.; Kan, H.; Koizumi, M.; Matsuda, A.; Shinji, S.; Yamagishi, A.; Yokoyama, Y.; Tatsuguchi, A. Utility of KRAS mutation detection using circulating cell-free DNA from patients with colorectal cancer. *Cancer Sci.* **2016**, *107*, 936–943.

(61) Weber, G.; Haslam, N.; Whiteford, N.; Prügell-Bennett, A.; Essex, J. W.; Neylon, C. Thermal Equivalence of DNA Duplexes Without Melting Temperature Calculation. *Nat. Phys.* **2006**, *2*, 55–59.

(62) Zhang, Y.-L.; Zheng, W.-M.; Liu, J.-X.; Chen, Y. Z. Theory of DNA melting based on the Peyrard-Bishop model. *Phys. Rev. E* **1997**, *56*, 7100–7115.

(63) Weber, G.; Haslam, N.; Essex, J. W.; Neylon, C. Thermal Equivalence of DNA Duplexes for Probe Design. *J. Phys.: Condens. Matter* **2009**, *21*, No. 034106.

(64) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*; Cambridge University Press: Cambridge, 1988.

(65) Taskova, M.; Uhd, J.; Miotke, L.; Kubit, M.; Bell, J.; Ji, H. P.; Astakhova, K. Tandem oligonucleotide probe annealing and elongation to discriminate viral sequence. *Anal. Chem.* **2017**, *89*, 4363–4366.

(66) Taskova, M.; Astakhova, K. Fluorescent Oligonucleotides with Bis(prop-2-yn-1-yloxy)butane-1,3-diol Scaffold Rapidly Detect Disease-Associated Nucleic Acids. *Bioconjugate Chem.* **2019**, *30*, 3007–3012.

(67) Taskova, M.; Barducci, M. C.; Astakhova, K. Environmentally sensitive molecular probes reveal mutations and epigenetic 5-methyl cytosine in human oncogenes. *Org. Biomol. Chem.* **2017**, *15*, 5680–5684.

(68) Taskova, M.; Astakhova, K. Fluorescent Oligonucleotides with Bis (prop-2-yn-1-yloxy) butane-1, 3-diol Scaffold Rapidly Detect Disease-Associated Nucleic Acids. *Bioconjugate Chem.* **2019**, *30*, 3007–3012.

(69) Miotke, L.; Barducci, M.; Astakhova, K. Novel Signal-enhancing approaches for optical detection of nucleic acids—Going beyond target amplification. *Chemosensors* **2015**, *3*, 224–240.

(70) Owczarzy, R.; You, Y.; Moreira, B. G.; Manthey, J. A.; Huang, L.; Behlke, M. A.; Walder, J. A. Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures. *Biochemistry* **2004**, *43*, 3537–3554.

(71) Drukker, K.; Wu, G.; Schatz, G. C. Model simulations of DNA denaturation dynamics. *J. Chem. Phys.* **2001**, *114*, 579–590.

(72) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **2011**, *134*, No. 02B627.

(73) Peyrard, M.; Bishop, A. R. Statistical Mechanics of a Nonlinear Model for DNA denaturation. *Phys. Rev. Lett.* **1989**, *62*, 2755–2757.

(74) Cuesta-López, S.; Peyrard, M.; Graham, D. J. Model for DNA hairpin denaturation. *Eur. Phys. J. E* **2005**, *16*, 235–246.

(75) Ferreira, I.; D Amarante, T.; Weber, G. DNA terminal base pairs have weaker hydrogen bonds especially for AT under low salt concentration. *J. Chem. Phys.* **2015**, *143*, No. 175101.

(76) Ivanova, A.; Rösch, N. The structure of LNA:DNA hybrids from molecular dynamics simulations: the effect of locked nucleotides. *J. Phys. Chem. A* **2007**, *111*, 9307–9319.

(77) Vester, B.; Wengel, J. LNA (Locked Nucleic Acid) High-Affinity Targeting of Complementary RNA and DNA. *Biochemistry* **2004**, *43*, 13233–13241.

(78) Pande, V.; Nilsson, L. Insights into structure, dynamics and hydration of locked nucleic acid (LNA) strand-based duplexes from molecular dynamics simulations. *Nucleic Acids Res.* **2008**, *36*, 1508–1516.

(79) Leonard, G. A.; McAuley-Hecht, K.; Brown, T.; Hunter, W. N. Do C–H...O hydrogen bonds contribute to the stability of nucleic acid base pairs? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1995**, *51*, 136–139.

(80) Asensio, A.; Kobko, N.; Dannenberg, J. Cooperative hydrogen-bonding in adenine-thymine and guanine-cytosine base pairs. Density functional theory and Møller–Plesset molecular orbital study. *J. Phys. Chem. A* **2003**, *107*, 6441–6443.

(81) Grunenberg, J. Direct assessment of interresidue forces in Watson-Crick base pairs using theoretical compliance constants. *J. Am. Chem. Soc.* **2004**, *126*, 16310–16311.

(82) Srinivasadesikan, V.; Sahu, P. K.; Lee, S.-L. Spectroscopic probe on N–H...N, N–H...O and controversial C–H...O contact in A–T base pair: A DFT study. *Spectrochim. Acta, Part A* **2014**, *120*, 542–547.

(83) Maximiano, R. V.; Weber, G. Deoxyinosine Mismatch Parameters Calculated with a Mesoscopic Model Result in Uniform Hydrogen Bonding and Strongly Variable Stacking Interactions. *Chem. Phys. Lett.* **2015**, *631–632*, 87–91.

(84) Weber, G. Mesoscopic Model Parametrization of Hydrogen Bonds and Stacking Interactions of RNA from Melting Temperatures. *Nucleic Acids Res.* **2013**, *41*, No. e30.

(85) Weber, G. Optimization Method for Obtaining Nearest-Neighbour DNA Entropies and Enthalpies Directly from Melting Temperatures. *Bioinformatics* **2015**, *31*, 871–877.