MASTERARBEIT | MASTER'S THESIS

Titel | Title

Adapting the SARSeq method for quantification of DNA-methylation

verfasst von | submitted by

Boris Paunovic BSc.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 834
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie
Betreut von | Supervisor:	Dr. Ulrich Elling

Acknowledgements

I would first like to thank my thesis advisor Dr. Ulrich Elling for giving me the chance to work in his laboratory and allowing me to collect a lot of experience in his group and for his keen interest in me at every stage of my research. His daily inspirations, timely suggestions with kindness, enthusiasm, and high level of expertise have enabled me to complete my thesis. The door to Dr. Elling's office was always open whenever I ran into a trouble spot or had a question about my research or writing. This was always possible, even while having a lot of balancing to do between all his students and the meetings with the WHO and his important leadership in advising politics and researchers in Austria and the world during the Covid-19 pandemic.

I would also like to show my biggest gratitude to Dr. Olga Frank as my daily supervisor and second reader of my thesis. Without her passionate participation and input, as well as her very valuable comments on this thesis, it could not have been successfully conducted. She consistently allowed this thesis to be my own work but steered me in the right direction whenever she thought I needed it.

Finally, I want to thank all the important departments at the Vienna Biocenter Campus (IMBA, IMP, VBC6), such as the protein facility, the molecular biology facility as well as the bioinformaticians which led us through the analyses, for helping us in achieving our goals.

Abstract

Background: Over one million people are dying of lung cancer, every year. A reason is the late diagnosis and, in many cases, only occurring symptoms in an advanced state of disease where metastasis has potentially already taken place. In lung cancer tumor suppressor genes are showing very often a pattern of methylated cytosines in their promoter CpG-rich regions, called CpG islands. These changes can be found in early stages of lung cancer and therefore it is very important to have a sensitive and specific technique for mapping such DNA mutations. Ultimately, the aim is to develop a high throughput method using the Saliva analysis by RNA sequencing technique (SARSeq) for quantification of DNA methylation.

Methods: In order to have an enrichment of target DNA, a Hybridization Capturing protocol, where magnetic beads that carry a pre-immobilized probe using oligonucleotides, is introduced. These oligos are designed to capture a specific DNA strand which could have the methylation pattern known to exist in lung cancer specific regions. Hybridization efficiency is tested while using qPCR as a readout.
The aim is to introduce an alternative to bisulfite sequencing, called the TET-assisted pyridine borane sequencing method (TAPS), where methylated cytosines found in CpG-regions get a direct positive conversion, what makes the readout easier and more approachable. In the end, combining those two methods and performing Hybridization Capture, followed by a TAPS protocol on the beads and finally using Next-Generation Sequencing for analysis is the aim.

Results: We optimized the hybridization protocol using the lung cancer specific locus ZFP42 for various capturing experiments. The optimization led to key findings, in terms of DNA purity, hybridization condition, beads washing and introducing a robust elution strategy, all leading to a higher hybridization efficiency. In addition, we presented a new type of hybridization probe, where the 5'-end as well as the 3'-end of a single strand get captured in a more efficient way, instead of catching the target strand only on one end, which we call bridge-capturing. The addition of a pre-denaturation step using DMSO and Betaine before starting hybridization had a promising impact on the capturing results. We successfully mastered the multiplexing of hybridization and captured various loci in a single reaction, a significant achievement which brought us closer to the goal of combining it with the TAPS method. The TET enzyme efficiency was improved by adding Bovine Serum Albumin (BSA) to the oxidation reaction, where half of the TET concentration in combination with BSA was sufficient for same results as twice the amount of enzyme without the additive, leading to a conversion of methylated and well accessible CGs of 100%. We crafted to perform the TAPS protocol on the beads, even succeeding in doing so on multiplexed beads and receiving successful CpG conversions in lung cancer specific regions ZP42, HOXa9, HOXa7, Kras, TAC1.

Conclusion: Our findings and integration of TAPS to the preceding hybridization capture suggest that the method could potentially be used in the future for an early detection of lung cancer as a medical diagnosis.

Zusammenfassung

Hintergrund: Jedes Jahr sterben über eine Million Menschen an Lungenkrebs. Ein Grund dafür ist die späte Diagnose und das Auftreten von Symptomen in einem fortgeschrittenen Krankheitsstadium, in dem möglicherweise bereits Metastasen gebildet wurden. Bei Lungenkrebs zeigen Tumorsuppressorgene sehr häufig ein Muster von methylierten Cytosinen in ihren CpG-reichen Promotorregionen, den so genannten CpG-Inseln. Diese Veränderungen können im frühen Stadium des Lungenkrebses gefunden werden, weshalb es sehr wichtig ist, über eine sensitive und spezifische Methode zur Detektion solcher DNA-Mutationen zu verfügen. Ziel ist die Entwicklung einer Hochdurchsatzmethode zur Quantifizierung der DNA-Methylierung unter Verwendung der SARSeq-Methode.

Methoden: Um eine Anreicherung der Ziel-DNA zu erreichen, wird ein Hybridisierungsprotokoll eingesetzt, bei dem magnetische Beads, die eine zuvor immobilisierte Sonde mit Oligonukleotiden tragen, eingeführt werden. Diese Oligos sind so konzipiert, dass sie einen bestimmten DNA-Strang einfangen, der das Methylierungsmuster aufweisen könnte, von dem bekannt ist, dass es in lungenkrebsspezifischen Regionen existiert. Die Hybridisierungseffizienz wird unter Verwendung von qPCR als Ausleseverfahren getestet.
Ziel ist es, eine Alternative zur Bisulfit-Sequenzierung einzuführen, die sogenannte TET-unterstützte Pyridin-Boran-Sequenzierungsmethode (TAPS), bei der methylierte Cytosine in CpG-Regionen direkt positiv umgewandelt werden, was das Auslesen einfacher und zugänglicher macht. Ziel ist es, diese beiden Methoden zu kombinieren und Hybridization Capture durchzuführen, gefolgt von einem TAPS-Protokoll auf den Beads und schließlich Next-Generation Sequencing für die Analyse zu verwenden.

Ergebnisse: Wir haben das Hybridisierungsprotokoll unter Verwendung des lungenkrebsspezifischen Locus ZFP42 für verschiedene Capturing-Experimente optimiert. Die Optimierung führte zu wichtigen Erkenntnissen in Bezug auf die DNA-Reinheit, die Hybridisierungsbedingungen, das Waschen der Beads und die Einführung einer robusten Elutionsstrategie, die alle zu einer höheren Hybridisierungseffizienz führten. Darüber hinaus haben wir eine neue Art der Hybridisierungssonde vorgestellt, bei der sowohl das 5'-Ende als auch das 3'-Ende eines Einzelstrangs auf effizientere Weise eingefangen wird, anstatt den Zielstrang nur an einem Ende zu fangen, was wir als Bridge-Capturing bezeichnen. Das Hinzufügen eines Vordenaturierungsschritts mit DMSO und Betain vor Beginn der Hybridisierung hatte einen vielversprechenden Einfluss auf die Einfangergebnisse. Wir haben das Multiplexing der Hybridisierung erfolgreich gemeistert und verschiedene Loci in einer einzigen Reaktion eingefangen, ein bedeutender Erfolg, der uns dem Ziel der Kombination mit der TAPS-Methode näherbrachte. Die Effizienz des TET-Enzyms wurde durch die Zugabe von Rinderserumalbumin (BSA) zur Oxidationsreaktion verbessert, wobei die Hälfte der TET-Konzentration in Kombination mit BSA für die gleichen Ergebnisse ausreichte wie die doppelte Enzymmenge ohne den Zusatzstoff, was zu einer Umwandlung von methylierten und gut zugänglichen CGs von 100 % führte. Wir haben es geschafft, das TAPS-Protokoll mit den Beads durchzuführen, und es ist uns sogar gelungen, dies mit Multiplex-

Beads zu tun und erfolgreiche CpG-Umwandlungen in lungenkrebsspezifischen Regionen ZFP42, HOXa9, HOXa7, Kras und TAC1 zu erhalten.

Schlussfolgerung: Unsere Ergebnisse und die Integration von TAPS in das vorangegangene Hybridisierungsprotokoll lassen vermuten, dass die Methode in Zukunft möglicherweise zur Früherkennung von Lungenkrebs als medizinische Diagnose eingesetzt werden könnte.

Table of Content

1. Introduction

1.1 General introduction

Lung cancer still causes over a million deaths every year around the world, which makes an early diagnosis and detection even more crucial, nowadays. Although there are new and modern medical techniques, as well as existing therapies, people are not going to the doctor as often as they should, or even not making regular appointments for a medical check, at all, which leads to the fact that such a disease often gets found at a stage where it is already far developed.

In order to be able to diagnose lung cancer, nowadays you have to make an appointment for a CT screening, which might have to be a few months in advance caused by the grand demand, as well as the limited personnel or screening devices. Furthermore, you would have to visit a hospital or another facility of choice, where people already get scared only thinking about it days or at least the night before. Not to mention the fact that for such a screening you would have to get stung because for a correct diagnosis the use of a specific contrast medium is required, in order to be able to differ cancer tissue from healthy ones and this is linked to additional fears and nervousness, not to speak of real anxiety and needle phobia. Additionally, with a CT screening it is only possible to diagnose such a disease, meaning nodules, at an already specific size and therefore specific stadium of mentioned disease.

There are several papers which got published that are going to be mentioned throughout this thesis and used together in a way that allows you to skip all the difficulties mentioned above, in order to try and develop a non-invasive method for a successful diagnosis at a still early stage of lung cancer.

Throughout the Covid-19 pandemic, the Austrian population got very used to the '*Alles gurgelt*' tactics. Gargling is a non-invasive way to catch genomic DNA (gDNA) from your lungs and eventually, the aim of this work is to help the development of a method where a patient takes a gargle test at home and hands or sends it in to his doctor of choice, without needing an appointment. Afterwards, it gets sent to a laboratory where the specific changes in your DNA could be screened.

There are tests on the market where blood samples are taken and tested for DNA methylation, which again is dependent on medical personnel as well as getting stung in order to get the blood from the patient. The finally developed method would be for consideration because the genes that we are looking for get enriched and it should target a cheaper, easier and quicker, as well as more precise way of developing a diagnosis targeting lung cancer while testing directly the genetic information coming from the screened organ.

1.2 Methylome

Looking at chromatin, there are histone tails which are targets of modifying molecules. The impact of the epigenetic modification we are looking at, which is DNA methylation, is in our case linked to silencing of specific DNA regions, especially promoters. It is an addition of a methyl-group to the cytosine residue in DNA, usually this occurs when the cytosine is next to a guanine residue. DNA methylation is usually found on both strands of DNA. There are, for example, some promoters that contain a very high level of CGs, which are called CpG islands. In cancer, tumor suppressors very often have mutations in the CpG sites, for instance, if we look at a promoter region of a tumor suppressor, it gets methylated and therefore silenced, which prevents the gene from properly performing its function against the formation of malignant cells.

1.3 Detection of methylated cytosines

In this chapter we will talk about how to detect methylated CpGs. Here we will compare two different strategies.

1.3.1 Bisulfite sequencing

The gold standard has always been the bisulfite modification. This is a handy tool because it always allows you to get information of methylated cytosines at a very high resolution. After the bisulfite protocol every unmethylated cytosine in a DNA region is converted to uracil. The chemical reaction does not work on methylated cytosines. Afterwards, if a PCR reaction is performed, we will retain the methylated cytosines as CG but the unmethylated cytosines will then be converted to TG. If then sequencing of these regions is done, we know that if we read a CG we had initially a methylated cytosine, if we find a TG we had initially an unmethylated cytosine. With bisulfite conversion you have to consider one thing: not only CGs are affected by the deamination reaction, but every C in the DNA region, even if it is not next to a Guanine. This method is very harsh, leads to DNA degradation and has this undirect readout after treatment, that is why we want to look at another technique for mapping DNA methylation which might be more approachable.

1.3.2 TET-assisted pyridine borane sequencing (TAPS)

Here we look at a bisulfite-free and direct detection of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC). It is called TET-assisted pyridine borane sequencing (TAPS)

where we get a positive readout for our searched methylated CGs. TAPS uses a ten-eleven translocation (TET) oxidation reaction of 5mC and 5hmC to 5-carboxylcytosine (5caC), followed by a pyridine borane reduction reaction to dihydrouracil (DHU).
It is able to detect modifications directly with high sensitivity and specificity, furthermore, does not affect the unmethylated cytosines at all.

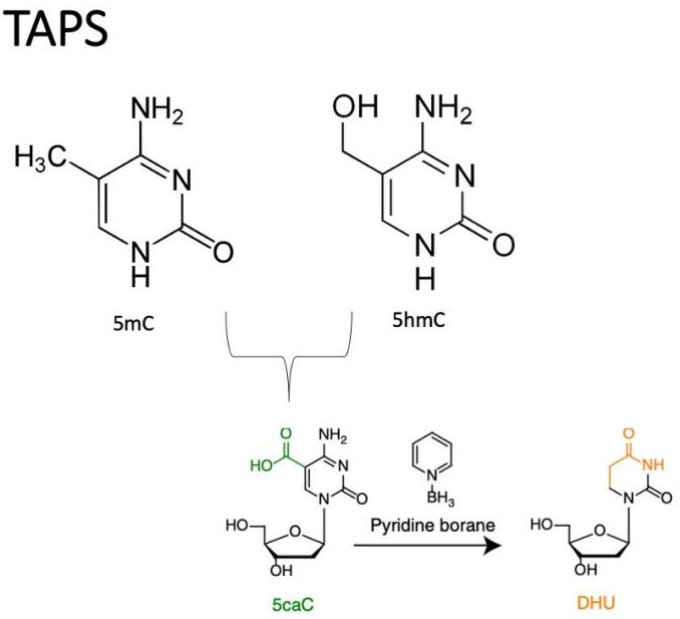

Figure 1 ***TAPS reaction scheme.*** *This picture depicts the reaction process from 5mC and 5hmC to DHU. First, the oxidation from 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) to 5-carboxylcytosine (5caC) takes place; followed by a pyridine borane reduction to dihydrouracil (DHU). (Picture adapted from [1]Liu, Y. et al. (2019)).*

Subsequently, after PCR the U becomes a T, means that our initial methylated CG is getting to a TG ([1] Liu, Y. et al. (2019)).

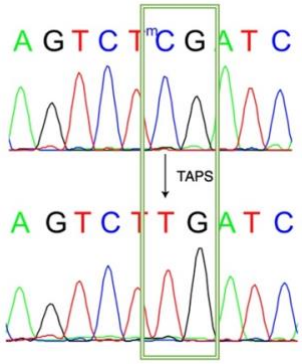

Figure 2 ***TAPS reaction after Sanger Sequencing.*** *After the conversion from C to U has taken place, and after PCR the uracil becomes a thymine; above is seen the initial sequence and on the bottom you can observe the readout after Sanger sequencing where the methylated CG gets converted to a TG. (Picture adapted from [1] Liu, Y. et al. (2019)).*

Depending on which strand you are looking at, we either get a conversion from CG to TG or to CA.

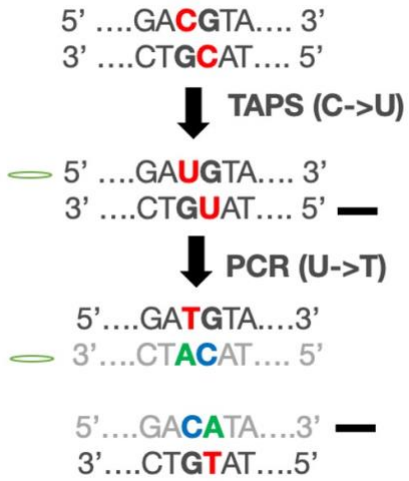

Figure 3 ***TAPS conversion model.*** *Depending on which strand (5'3' or 3'5') you are looking at, we either get a conversion from CG to TG or to CA.*

1.4 Lung cancer cells in sputum

With CT screens you can diagnose a patient with lung cancer. In a publication scientists used bisulfite sequencing for detection of promoter methylation by using sputum (and plasma) for following cancer-specific genes: ZFP42, HOXa9, HOXa7, TAC1, SOX17, CDO1 with a high sensitivity and specificity. These regions will later be considered for our experiments because the paper claimed that they have led to a high diagnostic accuracy for early-stage lung cancer that can be obtained using methylated promoter detection in sputum. ([2] Hulbert, A. et al. (2017)).

1.5 Hybridization Capturing

In the below mentioned paper, scientists performed a hybridization capture technique with urine cell-free DNA where they used oligos as probes that are pre-immobilized on magnetic beads and able to capture a complementary DNA region of interest. After putting it on a magnet, everything else in the supernatant will get washed out, means you get an enrichment of your target region. The idea is to process an initially large volume of up to 10mL and put it into one PCR well, in the end ([3] Oreskovic, A. & Lutz, B. R. et al. (2021)).

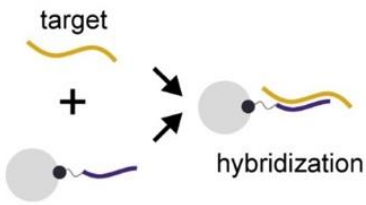

Figure 4 ***Concept of hybridization.*** *Target DNA strand gets captured by the probe which is bound to the magnetic bead, in order to get an enrichment of your region of interest, while getting rid of the other DNA that is not needed. (Picture adapted from [3] Oreskovic, A. & Lutz, B. R. et al. (2021)).*

1.6 SARSeq

The SARseq method offers a unique two-dimensional indexing strategy, which allows for well- and sample-specific labeling; means that you can pool all given samples and still every single sample can get analyzed on its own because it gets a specific barcode. The first dimension stands for a 96-well plate where every well (sample) gets a barcode, this whole plate gets pooled. In the end called plate gets a place in one well of a new 96-well plate. Finally, every well which is now representing one plate gets again one specific barcode, providing you with 96x96 possibilities ([4] Özkan, E. et al. (2021)).

2. Aim

The primary objective of this thesis was to combine the mentioned proof-of-concept papers, go away from the harsh bisulfite treatment, in using the TAPS method for an easier and direct positive detection of methylated CGs. In order to evade the 100ng DNA per TAPS reaction limitation, we are introducing a hybridization capture technique for an enrichment of target DNA and ultimately performing the TAPS protocol on the beads with the pre-captured strands of interest on it, which in combination was never done before. Therefore, we are aiming to develop a more sensitive method that will be able to map lung cancer specific methylation for an early detection of disease.
The idea is to introduce a quicker protocol and at the same a cheaper method, which is the reason for aiming a multiplexing of beads and primers for subsequent PCRs because NGS can be quite expensive but together with the SARSeq method it is already proven that such costs can be reduced due to using molecular barcodes and pooling. Finally, before combining the various ideas and methods, we sought to optimize and try to improve the given protocols in order to have a better approach on how to manage the aimed task of combining them and gain a better understanding of how to offer our help in developing a potential new strategy for medical diagnostical purposes.

3. Materials and methods

3.1 Materials

3.1.1 Chemicals, solutions, buffers, enzymes

TAPS Buffer:

The buffer was prepared according to the published method ([1]):

167mM Hepes (Sigma-Aldrich, H4034-100G) pH=8, 333mM NaCl (in-house), 3.3mM Alpha-Ketoglutarat (K3752, 5G Sigma), 6.67mM L-Ascorbic Acid (95210, Sigma), 4mM ATP (A6419, Sigma), 8,33mM 1.4-Dithiothreitol (DTT) (Roche, 10708984001). 5M NaOH (in-house), Mono Q H_2O (in-house), DEPC H_2O (Thermo Fisher Scientific), 1.5mM Ammonium iron (II) sulfate hexahydrate (Sigma-Aldrich, 09719-50G), Proteinase K (Sigma-Aldrich, 1.24568.0500). Albumin Bovine Serum (BSA) (VWR Prolab Chemicals, 422351S), Wash buffer (in-house), 5% Triton (in-house), DNA Purification MiniPex Kit (in-house), Zymo Kit (Zymo Research: EZ-96 DNA Methylation MagPrep). PCR reagents: TAQ enzyme (Hot Start & No hot start), 10x

Wash buffer used in TAPS on the beads protocol:

Stock	final conc
1M Tris pH8 (in-house)	20mM Tris
0.5M EDTA (in-house)	10mM EDTA
5% Triton (in-house)	0.2% Triton
	H2O

Table 1 Wash buffer. Ingredients and concentrations.

Hybridization buffer was prepared according to the published protocol ([2])

Stock	final conc
0.5M EDTA	20mM
1M Tris pH8	8mM
5 M NaCl	1M
10% Tween-20	0,10%

Table 2 Hybridization buffer. Ingredients and concentration.

TAQ-Buffer, dNTPS (all in-house; MolBio).

TET-enzyme (in-house, Protein Facility)

Dimethyl sulfoxide (DMSO) (Merck, 1.09678.0100)

5M Betaine solution (Alfa Aesar, J77507)

Restriction digest with enzyme NlaIII (NEB, R0125S).

Probes/Oligos for hybridization capture:

Probe_Zfp42_-_64C + Probe_Zfp42_-_5'_68C	GATTACACCCACGCGTATTTGTTCAACAGAC + GCTTCGCGGTAACAGGGGTGAGTCTTGTTT	"Bridge-Capturing"
Probe_ZFP42_-_40	GATTACACCCACGCGTATTTGTTCAACAGACATTTATTGA	40bp
Probe_ZFP42_+_64C	5' CCCAAACGTAAAACAAAGGAGAAACTGCAGT 3'	+strand, ~26bp (long)
Probe_ZFP42_+_58C	5' CCCAAACGTAAAACAAAGGAGAAAC 3'	+strand, ~20bp (short)
Probe_ZFP42_-_64C	5' GATTACACCCACGCGTATTTGTTCAACAGAC 3'	-strand, ~26bp (long)
Probe_ZFP42_-_58C	5' CACGCGTATTTGTTCAACAGACA 3'	-strand, ~20bp (short)
Probe_ZBTB12_-_64C	5' TGTGTGCATCTGCTCAGCCTCCCA 3'	-strand, ~26bp (long)
Probe_ZBTB12_-_58C	5' GATGTGTGCATCTGCTCAGC 3'	-strand, ~20bp (short)
Probe_RHBDF1_+_64C	5'CGGAGGTGACAGAGGTGATACTTGCTG3'	+strand, ~26bp (long)
Probe_RHBDF1_+_58C	5' GGTGACAGAGGTGATACTTGCTG 3'	+strand, ~20bp (short)
Probe_GAK_-_65C	5' GGCTTGGTGGGTCTTGACTGGG 3'	-strand, ~26bp (long)
Probe_GAK-_58C	5' GTCTTGACTGGGTGCCCT 3'	-strand, ~20bp (short)
Probe_GAPDH_2-_64C	5'GGGAAGGTGAAGGTCGGAGTCAACG3'	-strand, ~26bp (long)
Probe_GAPDH_2-_58C	5'GGAAGGTGAAGGTCGGAGTC3'	-strand, ~20bp (short)

Table 3 Probes for hybridization capture. Used for optimization experiments.

Probe_Zfp42_-_64C_3' Probe_Zfp42_-_5'_68C	GATTACACCCACGCGTATTTGTTCAACAGAC GCTTCGCGGTAACAGGGGTGAGTCTTGTTT
TAC1-_62_3' TAC1-_63_5'	ATGGGCATCGACGAGTTACCG GGTTGGAGAATCTTTGGGACGCGAT
HOXA7-_60_3' HOXA7-_61_5'	GCTCAAATATGCGGCCAAAGAATCC GGACTTAGACTCGGATCCAGACGG
SOX17-_63_3' SOX17-_60_5'	TGGGCAAGTACGTCGATTCCAAGG TGTAGACCAGACCGCGACAG
CDO1+_62_3' CDO1+_63_5'	CAGCCATCTCCTCCGACCCTT TCTTTGCCGGCGATGAGGTCA
GAK+_64_3' GAK+_63_5'	CCACACGTGAAAGGCCCAGGATG CCCAGTCAAGACCCACCAAGCC
HOXA9-_60_3' HOXA9-_59_5'	CACAATTAATCACGCCATCAAGAAGG AGGTTTTGAGGGCCTGGTTG
KRAS-57_3'	GTATCAAAGAATGGTCCTGCAC
GAPDH-_63_3' GAPDH-_61_5'	CTCCGGGTCTTTGCAGTCGTATGG GGCCCCGGGATGCTAGTG

Table 4 Bridge-Capturing Probes for hybridization capture.

gDNA Extraction from A549 lung cancer cells and oral keratinocytes:

LB -SDS

	Stock	dilution
10mM Tris (pH8)	1M	100
10mM EDTA (pH8)	0.5M	50
100mM NaCl	5M	50

VOL (mL)	500
10mM Tris (pH8)	5
10mM EDTA (pH8)	10
100mM NaCl	10
MilliQ (in-house)	475

LB +SDS

	Stock	dilution
10mM Tris (pH8)	1M	100
10mM EDTA (pH8)	0.5M	50
100mM NaCl	5M	50
2% SDS (in-house)	10%	5

VOL (mL)	500
10mM Tris (pH8)	5
10mM EDTA (pH8)	10
100mM NaCl	10
2% SDS	100
MilliQ (in-house)	375

Table 5 Extraction buffers. Materials for LB buffer with and without SDS.

3.2 Molecular biological methods

3.2.1 gDNA cell-extraction

The cell pellet was around 0.5mL. We add 10mL Lysis buffer (LB) to the cell pellet, as following:

5.5mL LB -SDS Buffer is added to the pellet and vortexed, after we mix it with 5.5mL LB +SDS Buffer. Next, 10% of total volume Proteinase K (1100µL) is added and the incubation needs to take place at 55°C overnight. On the next day we add in a ratio of 1:200 PureLink RNase A (20mg/mL), which is 60.5µL, and incubate it for 1 hour at 37°C. 12.2mL Isopropanol is mixed with the 12.2mL cell lysate and after adding 10% of total volume 3M NaAc (= 2440µL) we shake it until you see a DNA cloud. A centrifugation step is added for 10-15 minutes at full speed. We wash 2x with 70% EtOH and always spin down in between. After the second wash, we transfer the samples to a 2mL Eppendorf tube and dry the pellet, but not overdry it. If the pellet is very big you try to dry it in the heating block at a temperature of between 37-50°C until the edge of the pellet turn translucent. In the end, we add ~50µL H_2O and keep it in the heating block for 10 minutes in order for the EtOH to evaporate, but H_2O will not. Finally, we add more H_2O until the sample is not too viscous anymore.

3.2.2 NlaIII restriction digest

Figure 5 ***NlaIII restriction enzyme-cutting sites.***

*(https://www.neb.com/en/products/r0125-nlaiii#Product%20Information**; 01.10.23)***

	1rxn
gDNA A549	10µg
10x CutSmartBuffer	10µL
H_2O	Fill up to 100µL
NlaIII	1µL
Sum	100µL

Table 6 NlaIII restriction digest reaction mix.

.

The DNA needs to get digested in order to get appropriate lengths of strands, because the extracted DNA alone is very big and long, it needs to get smaller for the TAPS in solution experiments. The digest mix is incubated at 37°C over night; inactivation takes place at 65°C for 20 minutes.

3.2.3 Gargle preparation

Gargle with 4mL water and add it to 1ml 50mM EDTA and 50mM Tris-HCl (end conc 10mM EDTA and 10mM Tris-HCl).
Add 1/100 Volume 1M DTT (50µL; end conc 10mM) and incubate for 10 minutes at room temperature. Add 1/50 Volume 5%Triton (100µL; end conc 0,1%). Add 1/50 Volume 100µL ProK (100µL; end conc 2%). Incubate at 55°C overnight, inactivate at 75°C for 1 hour.
Prepare a 10x CutSmart buffer with 200mM $MgCl_2$:
1/9 volume of 10x CutSmart buffer (total volume 580µL) + 1/5 volume of 1M $MgCl_2$ (200mM): 464µL 10x CutSmart + 116µL 1M $MgCl_2$. Add 580µL 10x CutSmartBuffer + 0.2M MgCl2 to 5250µL Gargle. Add 1µL of NlaIII. Incubate at 37°C overnight, inactivate at 65°C for 20 minutes, spin down at maximum speed for 5minutes. In the end take 700µL of the supernatant for the hybridization experiment.

3.2.4 Hybridization & qPCR

The protocol was taken from the publication ([2]):

Wash the beads with H_2O: Vortex beads to ensure that beads are evenly dispersed in solution. Pipette 10uL beads into a 1.5mL Eppendorf DNA LoBind tube. Place beads on magnetic rack for 1 minute. Wash beads two times with 200µL H_2O. Place beads on magnetic rack for 1 minute. Remove and discard supernatant by pipette. Add equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20) and resuspend the beads.

Sample preparation:

Prepare **10µg digested** DNA per condition in 500µl H2O, add 50µl 0.5M EDTA and 10µl 1M Tris pH 8 and 440µL H_2O.

Capture gDNA by hybridization:

Add 250µL 5 M NaCl (final concentration 1 M), 12.6µL 10% Tween-20 (final concentration 0.1%) and 10µL prepared beads to each sample. Vortex beads well to ensure that they are evenly resuspended before adding to sample. Mix well by inversion. Denature for 15 minutes in a heating block at 95°C. Put samples for 1 hour on the rotator (slowest level of rotation, level 1). Cool samples down: put tubes in a cold block and let them rest in the cold room for ~5 minutes.

The wash step from the protocol ([2]) indicates to place the samples on a magnetic rack for 1 minute. We then remove the supernatant by pipette and discard it. You remove the tube from magnetic rack. Next, we add 0.5mL high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20) and wash it by inverting 10-20 times, or until no bead aggregate is left on the tube wall. We do not vortex but spin it down briefly. We place it again on the magnetic rack for 1 minute. After removing and discarding the supernatant by pipette the tube is removed again from the magnetic rack. 0.5mL high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20) are added and a washing step is performed by inverting the tubes 10-20 times, or until no bead aggregate is left on tube wall. We do not vortex but spin it down briefly. The tube is place on the magnetic rack for 1 minute. Afterwards, the supernatant is removed and discarded by pipette. After removing the tube from our magnetic rack, 0.5mL low salt wash buffer (15mM NaCl, 10mM Tris-HCl pH8, no Tween-20) is added and a washing step is performed by inverting 10-20 times, or until no bead aggregate is left on tube wall. We do not vortex but spin it down briefly. The tube is place on the magnetic rack for 1 minute. Afterwards, the supernatant is removed and discarded by pipette. Spin the tube down again, place it on the magnetic rack and remove as much liquid as possible.

Elute purified gDNA:

Prepare fresh 20mM NaOH by adding 990µL water to 10µL 2M NaOH. Add 20µL freshly prepared 20mM NaOH to beads to elute purified gDNA. Vortex for 5 seconds. Spin down

briefly. Place on magnetic rack. Transfer as much supernatant as possible (usually 20 – 21µL; set P20 pipette to maximum volume) directly to PCR well or to new Eppendorf DNA LoBind tube. This contains purified gDNA.
Partially neutralize with 3.5µL 100 mM HCl (prepare: 10µL 2M HCl + 190µL water).

Master Mix	
Component Volume Final Concentration	1 rxn
GoTaq. qPCR Master Mix (2x)	10µL
FW Primer (10µM) (0.5µM)	0.4µL
REV Primer (10µM) (0.5µM)	0.4µL
DNA	10µL
Sum	20.8µL

Table 7 qPCR Master mix.

Readout - Quantify by qPCR:
Application – Bio-Rad CFX Maestro → Program:
Initial denaturation: 95°C, 120 seconds
Denaturation: 95°C, 15 seconds
Annealing: 67.4°C, 30 seconds
Extension: 72°C, 30 seconds

The GoTaq qPCR master mix used BRYT Green as a fluorescent dye, which gets incorporated into double stranded DNA as it forms and due to this an increase in fluorescence is measured.

Calculation model 1.0:

We take digested and purified gDNA + NaOH/HCl Buffer (20mM) (same is used for elution + neutralisation):
5µL purified gDNA (25ng/µL) + 10µL 50mM NaOH + 10µL 50mM HCl → this equals 20mM NaOH and 20mM HCl + 125ng gDNA in 25µL = 5ng/µL → We are putting 50ng in the 10µL qPCR reaction.
According to the protocol, we were eluting in a volume of 20µL, 10µL are used for the qPCR, the other half was frozen.

Calculation model 2.0:

The difference here is that you take the input after the pre-denaturation step with DMSO and Betaine before you add the hybridization buffer. We dilute it 1:10 for qPCR (practically, 1:100) and put 10µL of it into the qPCR reaction.

In both cases the Ct-values from samples and inputs after the qPCR are compared and used for the calculation of the captured percentage from the initial gDNA that was used for the hybridization experiment.

3.2.5 TAPS in solution

The protocol was taken from the publication ([1])

Preparation of the Fe-Solution:

We prepare 30mL of nuclease-free water with 0,178g (=178mg) of Ammonium iron(II) sulfate hexahydrate to get a 15mM Fe-solution; we dilute it 1:10 to get the 1,5 mM Fe-solution.

Set up the oxidation reaction as follows (on ice):

DNA	depends on concentration	(Final amount up to 100ng)
TET oxidation buffer	15µL	(1x)
1,5 mM Fe	3,33µL	100µM
TET	depends on the batch and concentration	4µM
Nuclease-free water	(various: fill up to 50µL)	
Total	50µL	

Table 8 TET oxidation buffer mix.

The incubation reaction takes place at 37°C for 80 minutes.

After incubation 1µL (for 1rxn) Proteinase K (0,3 U) is added to the oxidation reaction and incubated for 1 hour at 50°C. PCR Purification in 35µL nuclease-free water is performed (Binding step: 1min/7000rfc; 2x wash + 2min dry).

The pyridine borane reduction is performed as followed:

We add 10µL of 3M sodium acetate solution pH=4.3, and 5µL 10M pyridine borane (toxin-room) to 35µL DNA sample, then we vortex. Incubation takes place at 37°C.

After 16 hours we perform a PCR Purification (30µL EB) and perform PCR1:
The #1PCR is with two primers (fw + rev) for amplifying our TAPS product. #2PCR is for adding the barcodes; thermocycler from BioRad is used for all reactions.

PCR1:
PCR program:
Lid: 105°C
Volume: 50µL

1. 95°C, 30s
2. 95°C, 15s
3. 68°C, 30s; -1°C per cycle
4. 72°C, 40s
5. GOTO step 2, 10x
6. 95°C, 15s
7. 58°C, 30s
8. 72°C, 40s
9. GOTO step 6, 30x
10. 72°C, 3min
11. 12°C, cool down

Table 9 PCR1 program.

Master Mix	1 rxn
H2O	40,50µL
PCR2 10xBuffer	5µL
dNTPS (25mM)	0,5µL
10uM fw Primer	1µL
10uM rev Primer	1µL
DNA (5-10ng)	1µL
TAQ	1µL
Sum	50µL

Table 10 TAPS PCR1 master mix.

Perform EXOProStar reaction in order to get rid of the primers, instead of a column purification:
2µL EXOProStar enzyme + 5µL of every PCR1 product; (in total 7µL).
→ EXOStar-Program (Thermocycler):
Incubation at 37°C for 30 minutes,
Inactivation at 80°C for 15 minutes
Hold at 12°C.

PCR2-Program:
Initial denaturation: 95°C, 3 minutes
Denaturation: 95°C, 30 seconds
Annealing: 65°C, 30 seconds
Extension: 72°C, 40 seconds
for 7 cycles,
Hold: 72°C, 3 minutes then 4°C for cooling down.

After EXOStar: PCR2 for adding the EB-Barcodes

PCR2 MM	1rxn
H2O	33,5µL
Buffer	5µL
dNTPs	0,5µL
TAQ	1µL
Sum	40µL
DNA	2,5µL
EB	7,5µL

Table 11 TAPS PCR2 master mix.

After PCR2:
Pool the samples and put it on a Gel. Perform Gel Extraction (Mini-Pex Kit) and NGS.

3.2.6 Combining Hybridization Capture & TAPS on the beads

After Hybridization we wash the beads with 1mL 70%-EtOH and dry them. TET enzyme is only able to perform its function on double stranded DNA, and after hybridization we have only

single strands captured. In order to fill up the seconds strand, we introduce a second strand synthesis, also “extension on the beads”. We add an extension-mix where we put H_2O, 10x TAQ Buffer, dNTPs, and TAQ enzyme (Not Hot-start!).

Extension on the beads:

Keep the samples on ice and add the extension-mix to the beads:

1x TAQ Master Mix	1 rxn
H_2O	43.5µL
10x TAQ Buffer	5µL
dNTPS (25mM)	0,5µL
TAQ (Not Hot-start!)	1µL
Sum	50µL

Table 12 Second strand synthesis-mix. Used for the extension on the beads.

Prepare double the amount for every sample and add 100µL extension-mix to the beads, then separate it twice to 50µL. Put the samples in the heating block with an open lid (in order for residual EtOH to evaporate), turn on the heater and incubate at 60°C for 30 minutes (after first 5 minutes close the lid), then heat it to 72°C and incubate for 5 minutes. After the extension protocol is done, put the samples immediately on ice and put them on the magnetic rack. Wash the beads 2x with H_2O+Triton, spin them down and take out all the liquid.

Continue with TAPS:

After TET and Proteinase K (ProK) incubation:

Wash with 200µL Binding Buffer (Buffer B from MiniPex-Kit)

Wash with 200µL 100mM Tris+10mM EDTA

Wash 3x with 1mL H_2O.

After last wash add 35µL H_2O (per 1rxn) to the beads.

→ Set up the borane reaction:

After borane:

Wash with 3x 50µL Wash Buffer

Wash with 1x 50µL H_2O

Next, remove H_2O from the rest of the beads as much as possible (magnet), then add10µL H_2O to the beads, and use 2µL for the first strand specific linear amplification where we want to amplify only the target strand which gets captured by the beads and is initially methylated, using only one primer (either the one for the FW or the REV strand). The next step is the 1st SARSeq PCR where the linearly amplified single strands get now processed in a PCR where primers for both FW as well as REV are used in order to create double stranded DNA again.

The 2nd SARSeq PCR is used to incorporate barcodes into the DNA in order to be able to perform multiplexing of all the samples for later identification distinction of DNA samples during NGS analysis.

ssPCR (PCR1)

MM	0,5rxn
H2O	10,75
10x TAQ Buffer	2,5
dNTPs	0,25
Primer 0,5µM	9
Mixed Beads	2
TAQ	0,5
Sum	*25*

add 23µL MM

take 25µL supernatant for PCR (1.5) MM

PCR(1.5)	
MM	0,5rxn
H2O	3,25
10x TAQ Buffer	2,5
dNTPs	0,25
Primer FW 10µM	9
Primer REV 10µM	9
TAQ	1
Sum	*25*

Table 13 TAPS ssPCR(1) and PCR (1.5) master mix.

Perform EXOProStar reaction in order to get rid of the primers, instead of a column purification:
2µL EXOProStar enzyme + 5µL of every PCR1 product; (in total 7µL).
→ EXOStar-Program (Thermocycler):
Incubation at 37°C for 30 minutes,
Inactivation at 80°C for 15 minutes
Hold at 12°C.
PCR2-Program:
Initial denaturation: 95°C, 3 minutes
Denaturation: 95°C, 30 seconds

Annealing: 65°C, 30 seconds

Extension: 72°C, 40 seconds

for 7 cycles,

Hold: 72°C, 3 minutes then 4°C for cooling down.

After EXOStar: PCR2 for adding the EB-Barcodes

PCR2 MM	1rxn
H2O	33,5µL
Buffer	5µL
dNTPs	0,5µL
TAQ	1µL
Sum	40µL
DNA	2,5µL
EB	7,5µL

Table 14 TAPS PCR2 master mix.

After PCR2:

Pool the samples and put it on a Gel. Perform Gel Extraction (Mini-Pex Kit) and NGS.

3.2.7 Sanger Sequencing

In order to get quick results for a first overview, we perform the TAPS protocol until the EXOStar reaction is done, afterwards we dilute them and submit the samples for Sanger sequencing. For our PCR products which lay in the range between 200-500 bp, we submitted 3ng. In total, we submit 1µL of a 5µM primer together with 6µL of 0.5ng DNA.

3.2.8 Next Generation Sequencing (NGS)

After the TAPS protocol there is the first strand specific linear amplification where we want to amplify only the target strand which gets captured by the beads and is initially methylated, using only one primer (either the one for the FW or the REV strand). The next step is the 1st SARSeq PCR where the linearly amplified single strands get now processed in a PCR where primers for both FW as well as REV are used in order to create double stranded DNA again. The 2nd SARSeq PCR is used to incorporate barcodes into the DNA in order to be able to perform multiplexing of all the samples for later identification distinction of DNA samples during NGS analysis. After the 2nd PCR the product is loaded on a gel and gel extraction is performed. Finally, the purified samples are submitted for sequencing.

4. Results

4.1 Hybridization Capture Protocol Optimization

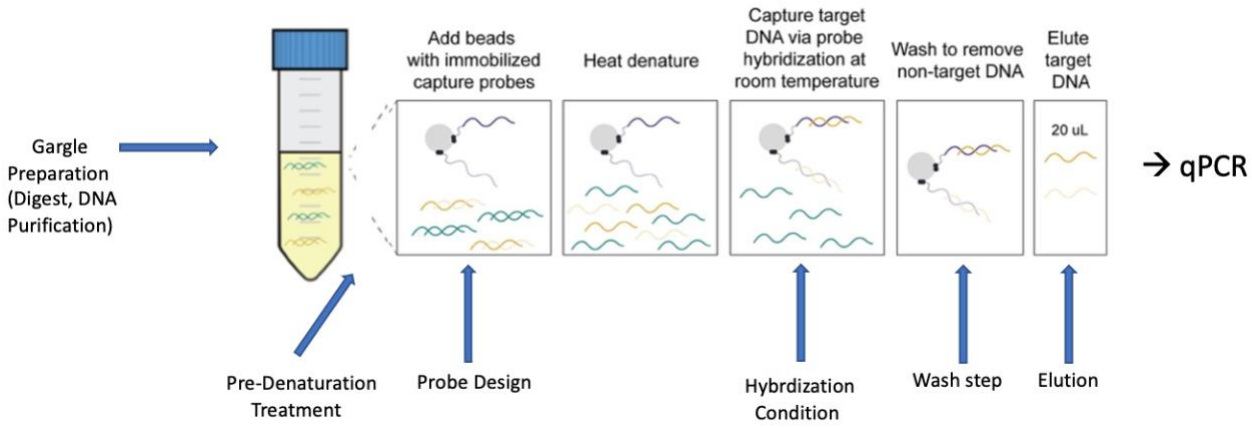

Figure 6 ***Hybridization Workflow.*** *The arrows indicate the protocol steps which were optimized in this study: Gargle preparation (Digest, DNA Purification); Pre-Denaturation treatment; Probe design; Hybridization Condition; Wash step; Elution. (Picture adapted from Oreskovic, A., et al. [3]).*

We optimized a previously published hybridization capture protocol, that utilizes magnetic beads with immobilized capture probes to selectively capture short cell-free DNA in urine samples ([3]). Our goal was to capture lung DNA in gargle samples, requiring the Proteinase K treatment, to remove histones and other proteins, thereby enabling the direct in-gargle DNA digest with NlaIII restriction enzyme (described in the method section). Subsequent DNA purification was performed to increase the capture efficiency (Figure 6, gargle preparation), the results are shown in chapter *4.1.7*. Our optimization efforts to increase the capture efficiency were conducted using DNA from A549 lung cancer cell line, with a focus on the following steps: DNA pre-denaturation treatment (chapter *4.1.3*), probe and bead design (chapter *4.1.2* and *4.1.4*, hybridization condition and wash step (*4.1.1*), (see Figure 6). Additionally, we modified the elution strategy (*4.1.1*), which was only used to validate the capture efficiency by qPCR (described in the method section).

This chapter will provide a detailed explanation and results regarding these specific points. For our optimization experiments we selected a lung cancer-specific hypermethylated region from the publication mentioned earlier in the introduction, known as Zinc Finger Protein 42 (ZFP42) ([2]). Each probe is designed to capture only one DNA strand. In this thesis '–' indicates the minus strand and '+' the plus strand. It is important to note that because only one DNA strand is captured, the theoretically maximum amount of captured DNA would be 50%. In the next paragraphs all experiments were done with the ZFP42- probe (probe that captures the minus strand of the ZFP promoter region).

4.1.1 Optimizing Hybridization Condition and Wash step with varying Elution strategies

First, we tested if the published protocol was optimal for our needs as we aimed to capture longer DNA fragments. The original method involves slow rotation (using a rotator) at room temperature for 1h. Under this condition, the beads with the immobilized probes remain in solution during each rotation, increasing the probability of capturing the dissolved target DNA. Our primary objective was to extend the time at higher temperatures while maintaining rotation, as we hypothesized that this would enhance the specific binding of long DNA fragments to the probes. To achieve this, we explored alternative conditions:

1. Hybridization at a constant temperature of 65°C with continuous rotation for 1 hour using a hybridization oven.
2. A slower cooling process by placing the samples under a Styrofoam box on top of a pre-heated metal block on a shaker immediately after the denaturation step at 95°C.
3. Controlled cooling down using a heating block, gradually reducing the temperature from 95°C to 65°C with a 30-minute incubation, then to 55°C with another 30-minute incubation, and finally cooling to room temperature.

In Figure 7A the results show the differences in the capture efficiency among different hybridization conditions. The original method (labeled as #4 in Figure 7A) indicates the best capturing efficiency, at 5.6% from input, while the alternative conditions showed either no hybridization (hybridization at constant temperature, #1 and slower cooling down (#2)) or only a minimal capture efficiency at 0.16% (controlled cooling down, #3). Consequently, we have chosen to proceed with the original method of hybridization at room temperature as the best condition.

The calculated capture efficiency of non-specific DNA (measured using the non-target region EEF1a) is very low (<= 0.002), indicating that we have already achieved specific binding and the removal of non-specific DNA, resulting in a substantial enrichment (see background and enrichment data in Figure 7A).

The low capture efficiency may not only be due to poor capture but also potentially harsh washing steps that could result in some loss. Therefore, in the next step, we examined whether using an 70% EtOH wash step instead of the high salt and low salt buffers (see method) would improve the capture efficiency while still effectively removing background.

As shown in Figure 7B, the use of 70% EtOH resulted in slightly higher capture efficiency (compared to first measurement in Fig. 7A), potentially removing less background but still enough to achieve over a 300-fold enrichment. Notably, this time the original method only showed a capture efficiency of 0.8%, significantly lower than our initial measurement. This discrepancy may be explained by the initial high variability introduced by harsh washing steps

or elution using NaOH (see methods), which can interfere with qPCR when the pH of the eluate isn't sufficiently neutralized.

To address this, we repeated the experiment by replacing the NaOH elution step with H_2O denaturation (described in the methods). The capture efficiency with the original protocol increased but remained lower at 2.26%, while the EtOH wash step yielded similar results of 6.8% to those in Figure 8B (see Figure 7C). The lower capture efficiency with the original protocol and elution using H_2O might still indicate the impact of harsh washing steps.

In conclusion, we have retained the hybridization method from the original protocol but made changes to the wash step and the elution strategy to calculate our % input, as we believe this approach is more robust.

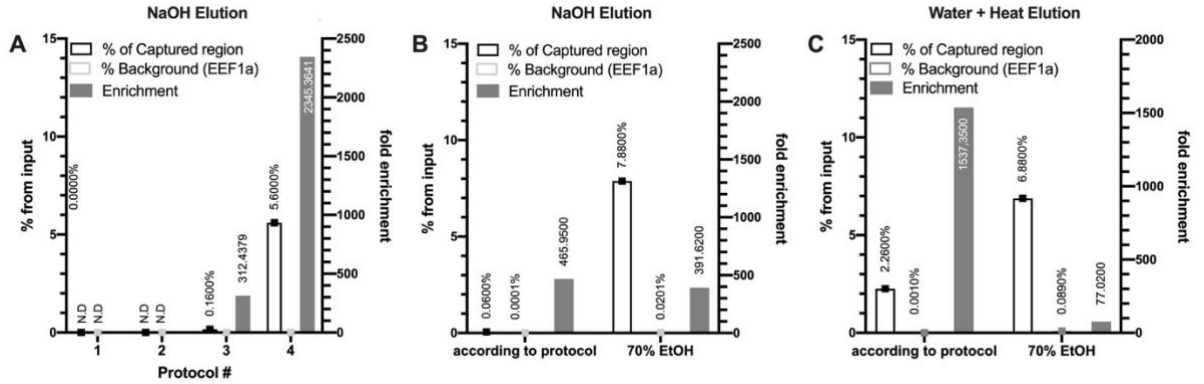

Figure 7 ***Optimization of hybridization condition and wash step with varying elution strategies.*** *The left y-axis shows the percentage of the captured region from the input (see methods for details). Right y-axis depicts the fold enrichment of the target region against the background, measured by non-captured region of EEF1a. NaOH Elution represents the standard elution method based on alkali denaturation, as described in the method section. Water + Heat elution refers to the newly introduced elution strategy, which applies the heat denaturation with water (for details see methods). A) Optimization of Hybridization conditions: Protocol# 1. Hybridization at constant temperature, #2. Slower cooling down, #3. Controlled cooling down, #4. original method (slow rotation at room temperature for 1h). The standard protocol with the rotor showed the best results. B) Comparison of washing steps between the original protocol and the washing step with 70% EtOH. The 70% EtOH washing step showed slightly improved results for capture efficiency compared to the measurements in Figure 8A and also indicates higher capturing results than the sample treated with the standard protocol where salt buffers are used for washing the beads. C) Comparison of washing steps (like in B)) under new elution strategy by denaturation: 1. Standard hybridization protocol and 2. Hybridization protocol with 70% EtOH washing step. The results for 2. shows ~3x higher capture efficiency than when washed standardly with salt buffers. N.D. indicates No Data*

4.1.2 Probe design

After optimizing the hybridization step, introducing a new wash as well as elution step, both the long and the short probes for the ZFP42 promoter regions are compared in an experiment where we studied the effect of oligo length on the capturing method. We wanted to see whether the longer oligo length of 26bp is resulting in a higher capture efficiency compared to the shorter ones at around 21bp. This time the probes capturing the minus strand as well as the ones for the plus strand were tested for comparison because the initial idea was to combine the beads with probes that capture the -strand as well as the +strand with the aim of getting a

higher capture efficiency when catching both strands of one locus. As shown in Figure 8, the probes designed for the +strands work less effectively for both long and short, with a capture of 1%-2%, respectively. This can be explained as with the +probe a different end of the target loci is captured which might behave differently. The ZFP long- probe used in this experiment has a capturing ability of 9%, whereas the ZFP short- probes show a slightly lower capture with 6%. According to these results, we will continue with the beads for the ZFP long- strands for our further optimization experiments because it showed the best results.

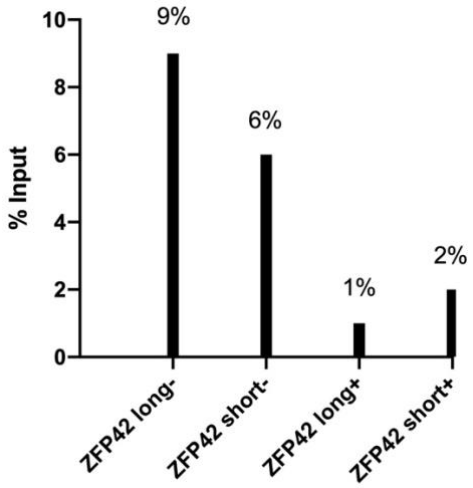

Figure 8 ***Comparison of probe length:****–On the y-axis there is the percentage that we capture normalized to the input; the x-axis depicts the probe design and region. The ZFP42 "long-"probe shows the highest capturing ability. In general, the probes for capturing the minus strand show better results.*

4.1.3 Pre-Denaturation treatment

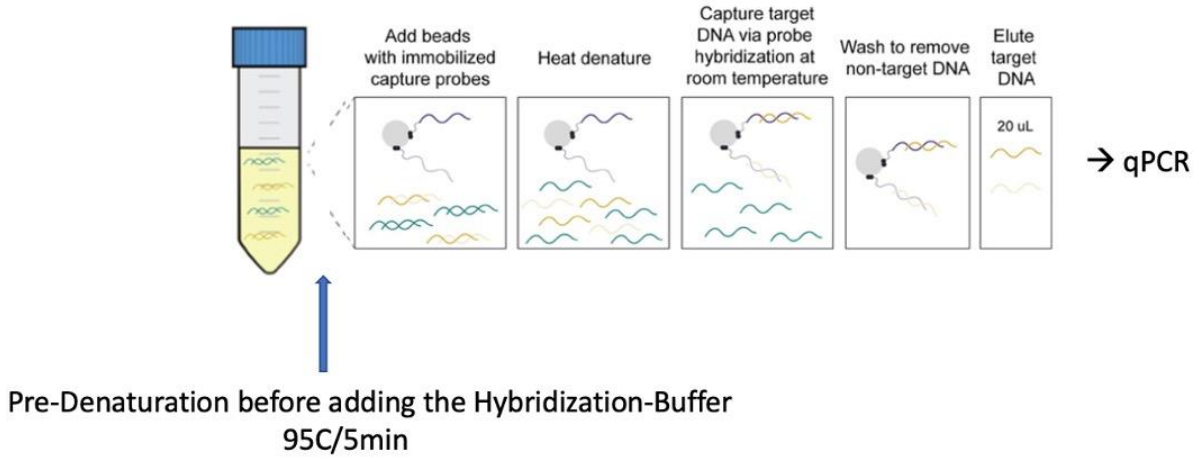

Figure 9 ***Introduction of a Pre-Denaturation step into the hybridization protocol.*** *(Picture adapted from [3].)*

The concept of a pre-denaturation step was devised with the idea that we have long CG-rich DNA regions and perhaps the denaturation step at 95°C might not be sufficient, which led us to the assumption that this additional step will help us in achieving a more effective

denaturation. Before the hybridization buffer and the beads are added to the DNA sample, we introduced a pre-denaturation step with DMSO and Betaine which is performed at 95°C for 5 minutes (see methods for details). We tested the pre-denaturation in three independent experiments (Figure 10B) and compared the results to the data from experiments without the pre-denaturation step (Figure 10A). To account for any potential positive effects of pre-denaturation on PCR itself, we took an aliquot from each sample (with and without pre-denaturation) before adding hybridization buffer and beads (for details see method section). The 1:10 dilution was used in qPCR for input measurements.

The ZFP42 long- probe which is used for capturing the -strand in Figure 10 shows a strong increase in capturing efficiency when a pre-denaturation treatment is introduced, being up to ~3x higher with 22.3% than without the denaturation step (7.67%). The background is with 0.17% slightly higher for the pre-denaturation treated sample, in comparison we have 0.09% background for the case where no pre-denaturation step is done, both still being very similar. The enrichment is ~30% higher with the pre-denaturation protocol with around 130, in contrast the standard condition without the additional pre-step results in an enrichment of ~100.

In conclusion, we will proceed with this new addition to the protocol, where a pre-denaturation step at 95°C for 5 minutes is introduced before adding the hybridization buffer and the magnetic beads to the DNA sample.

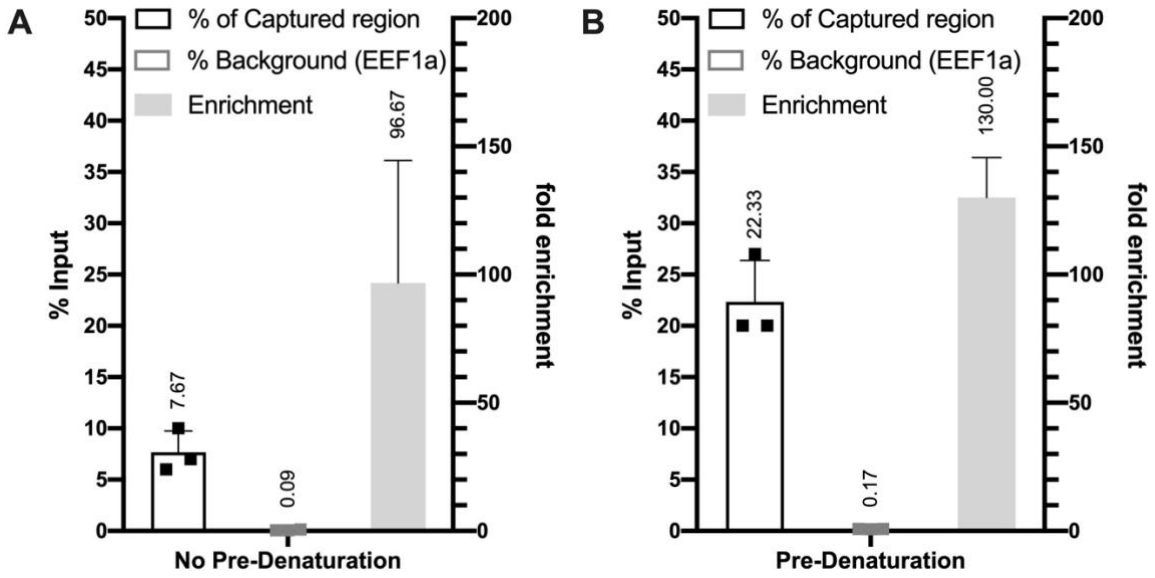

Figure 10 ***Pre-Denaturation treatment.*** *The left y-axis shows the percentage of the captured region from the input (see methods for details). Right y-axis depicts the fold enrichment of the target region against the background, measured by non-captured region of EEF1a. We deal with the comparison between the standard denaturation and an additional pre-denaturation step before adding the hybridization buffer and the beads to the DNA sample. A) No pre-denaturation, B) Pre-denaturation. The pre-denaturation step shows a 3x higher capturing efficiency. (n=3)*

4.1.4 Bead design

To increase the capture ability of the beads, we have tried a different design that enables capturing at two different locations within the same DNA region. This new type of bead, which we refer to as the “Bridge-Capturing bead” is equipped with two covalently bound probes: one for 3’ end and one for the 5’ end of the region of interest.

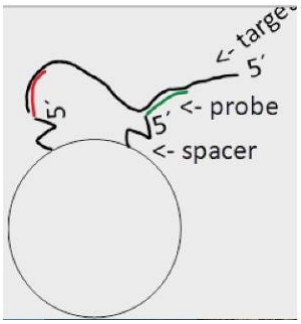

Figure 11 ***Illustration of the Bridge-Capturing Bead design.*** *Concept where both ends of one single strand are captured by one bead.*

In the following experiments where the new type of bead is introduced for ZFP42 minus strand region, we compared the linear-capturing ones with the bridge-capturing variant, in both cases DNA is treated with a pre-denaturation step with DMSO and Betaine.

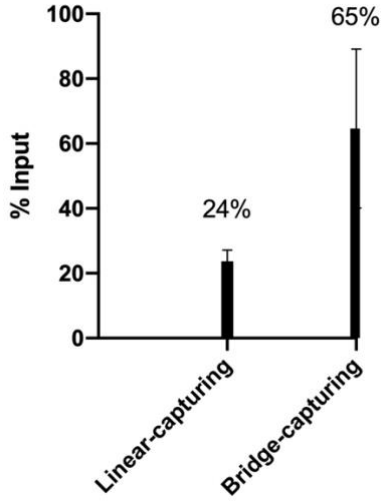

Figure 12 ***Comparison of the two bead designs****. On the y-axis there is the percentage that we capture normalized to the input; the x-axis depicts the probe design. The new bridge-capturing beads are ~3 times more efficient than the linear-capturing ones. (n=3)*

Figure 12 shows the capturing efficiency of the bridge-capturing bead is indeed higher than with the regular ones (24%). So far, the best result is reached with a 65% capture, normalized to the input.

In conclusion, after these results we decided to continue with the bridge-capturing type of beads for lung cancer specific regions from the publication Hulbert, A. et al. (2017) [2].

4.1.5 Test New Bridge-Capturing probes

After optimizing the hybridization capture protocol and determining that longer probes and a bridge-capture beads design are most effective for our purposes, we designed new probes for lung cancer-specific regions ZFP42, HOXa9, HOXa7, TAC1, SOX17 and CDO1, as described in the previously cited publication[2]. Additionally, probes for GAK were designed to serve as a positive control in later TAPS experiments, given its highly methylation levels in both healthy and cancerous cells. The GAPDH region was chosen as the negative control since it is consistently unmethylated. To further enhance the specificity of our assay, we included the KRAS region with the lung cancer-specific mutation at amino acid position G12, providing an additional indicator of the presence of lung cancer DNA in gargle samples. For loci CDO1 and GAK, probes capturing the + strand were also designed, while all other probes captured only the - strand of their specific regions.

The probes (GAK+ and CDO1+) for capturing the plus region of GAK and CDO1 worked much better than the probes for the -strand (GAK-, CDO1-), as indicated in Figure 13 where we have 35% in contrast to 8% for CDO1 and 23% in comparison to 16% for GAK. Figure 13 also presents the results showing ZFP42-, HOXa7- and GAPDH- were above the 50%-line, whereas other regions being able to accumulate around 25-35% of DNA from their input, Kras as well as HOXa9- being the weakest capturing probes in this experiment with 15% and 6%. The background is shown to be similar to previous experiments.

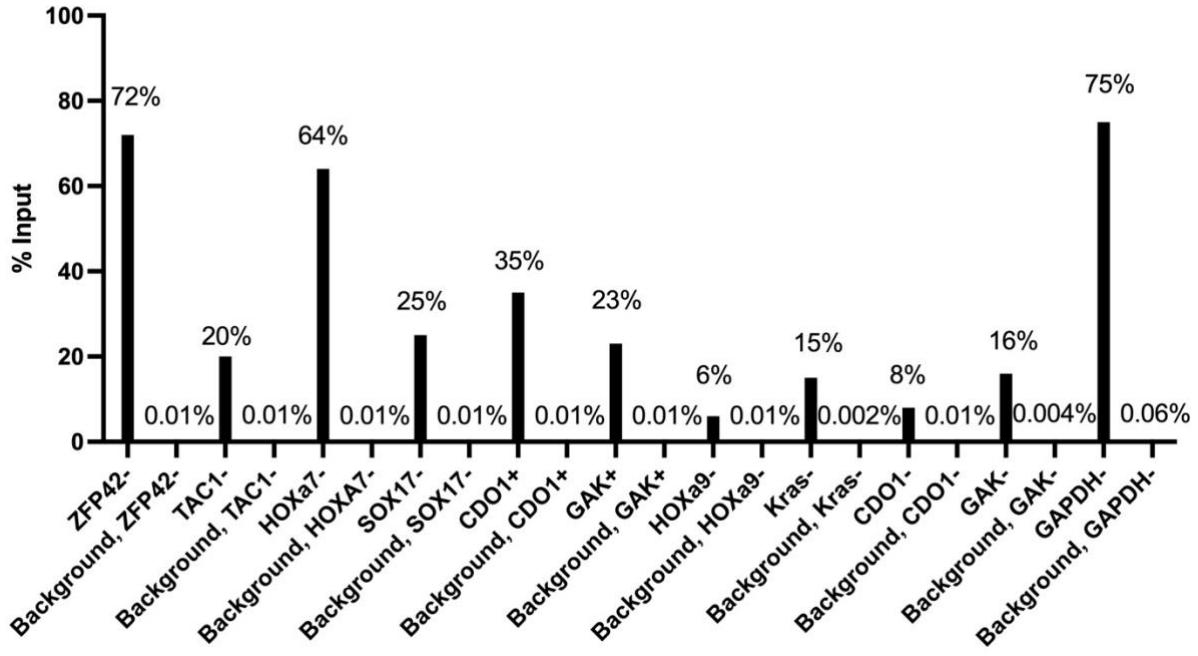

Figure 13 ***New bridge-capturing probes:*** *The x-axis represents different regions, each with its respective background indicated adjacent to it. The y-axis denotes the percentage of capture, normalized to the input. The bridge-capturing beads designed for the lung cancer specific regions work and show overall a good capturing efficiency. ZFP42- was once more the strongest competitor.*

After successfully performing hybridization with the new probes, a next step can be taken in performing the multiplexing of hybridization.

4.1.6 Multiplexing of Hybridization

As our goal is to establish a high-throughput detection method, the next step involved multiplexing all nine different probes (ZFP42-, TAC1-, HOXa7-, CDO1+, GAK+, HOXa9, KRAS-, GAPDH-)—which we previously tested (Figure 13)—in a single hybridization reaction, followed by TAPS and PCR reactions. Maintaining a total bead volume of 20 µL, we combined 2.22 µL for each of the nine target-specific beads in one reaction, and subsequently carried out our optimized hybridization protocol. Figure 14 shows that the multiplexing works, the percentages are lower than using 10µL (in Figure 13), except for KRAS- being the strongest with ~45% capturing efficiency, followed by ZFP42- with 30%. The average capturing percentage is around 6% for the other beads.

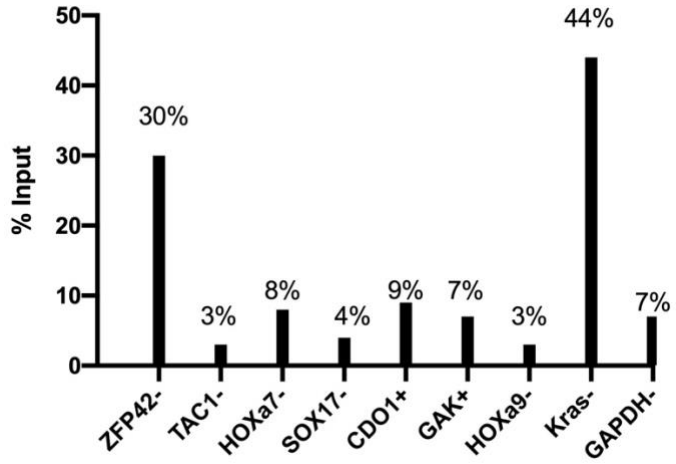

Figure 14 ***Multiplexing of hybridization.*** *The y-axis represents the percentage of captured DNA normalized to the input; the x-axis depicts the region. Multiplexing of hybridization by mixing all beads together in one reaction. 2.22µL of each bead was taken. The percentages of DNA capturing are lower than with 10µL per single bead, as assumed.*

In conclusion, the tests could show that multiplexing of hybridization is working for our protocol, according to our results beads for regions ZFP42- and KRAS- have the highest capturing ability of 30-45%, whereas the other beads in average deliver a percentage of captured DNA in a range of around 5-10% when initially mixing 2.22µL per bead together for one hybridization reaction.

4.1.7 Gargle: Testing different conditions - spin of samples, purification

All optimization experiments for the hybridization protocol were conducted using extracted and digested A549 cell DNA. Consequently, the next step involved testing our optimized hybridization protocol on gargle samples. Unlike samples from cell culture, gargle samples

contain various unwanted substances, including residual food or drink, mucus, and other contaminants from the patient, which could potentially reduce hybridization efficiency. To address this, we experimented with both short and long spin-down steps after the enzyme inactivation at 65°C and short and long spin-down steps after the inactivation step at 95°C. We hypothesized that the centrifugation step would eliminate interfering particles more effectively, particularly after the heating step, as it might lead to the coagulation of proteins.
The different inactivation- and spinning-conditions were chosen as followed:
65C for 20 minutes / spin down for 5 minutes
65C for 20 minutes / spin down for 15 minutes
95C for 10 minutes / spin down for 5 minutes
95C for 10 minutes / spin down for 15 minutes.
Additionally, one condition involved column purification (for details see methods) of the gargle DNA before performing the hybridization protocol.
For this experiment four regions (ZFP42-, HOXa7-, CDO1+, GAK+) were selected for the capturing process, and their respective beads mixed together with 5µL each to achieve a total volume of 20µL per reaction, consistent with the previous multiplexing experiments.
As indicated in Figure 15, an increased inactivation temperature in combination with a longer spinning time has a significant positive effect on hybridization capture efficiency particularly for ZFP42- probe. However, the most notable result is shown for purified gargle, with all regions exhibiting significantly higher capture efficiency compared to unpurified DNA under all tested conditions.

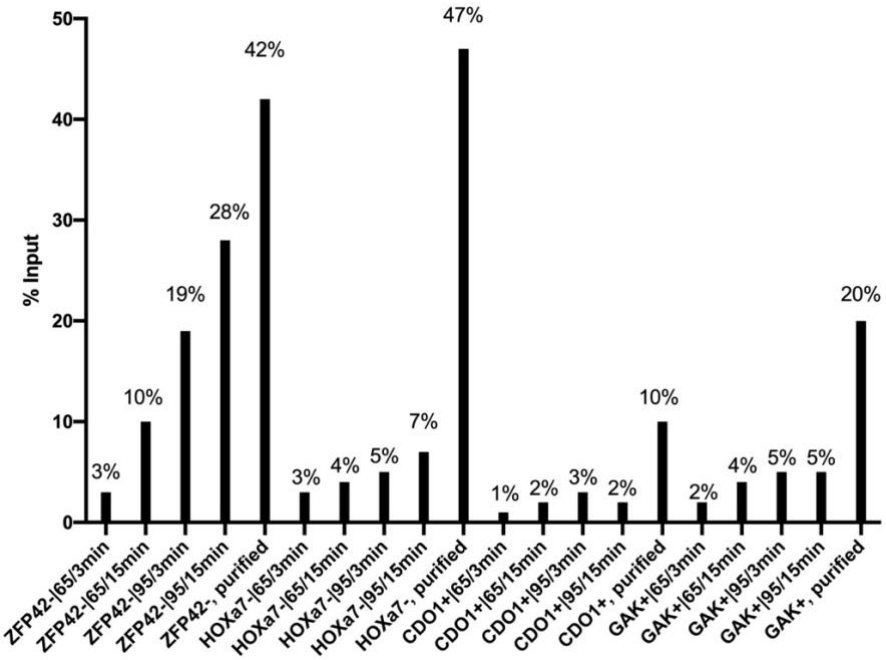

Figure 15 ***Gargle – Different conditions****. The y-axis represents the percentage of captured DNA normalized to the input, while the x-axis depicts the region, along with different conditions for each sample. The pattern observed indicates that higher inactivation temperatures and longer spinning times lead to increased capture efficiency. Notably, gargle purification shows a substantial difference in capture efficiency.*

Gargle: Purified vs. Unpurified

The conclusion after testing various gargle treatments was that purified gargle results in the highest hybridization capture efficiency. This is the reason why the next experiment was set up for comparison of digested gargle DNA which was either purified or not. This time all the probes were used again in a multiplexing manner as before where 2.22µL per bead was put into the mix. After the restriction digest the DNA was treated with the standard protocol concerning the inactivation temperature of 65ºC and spinning the solution down for 3 minutes. Figure 16 confirms the impact of purifying the gargle DNA before starting with the hybridization procedure. Unpurified samples seem to not be able to get captured because of potential disturbing particles and substances in the gargle solution. We will proceed with this strategy and test it again for A549 cancer DNA in the next experiment.

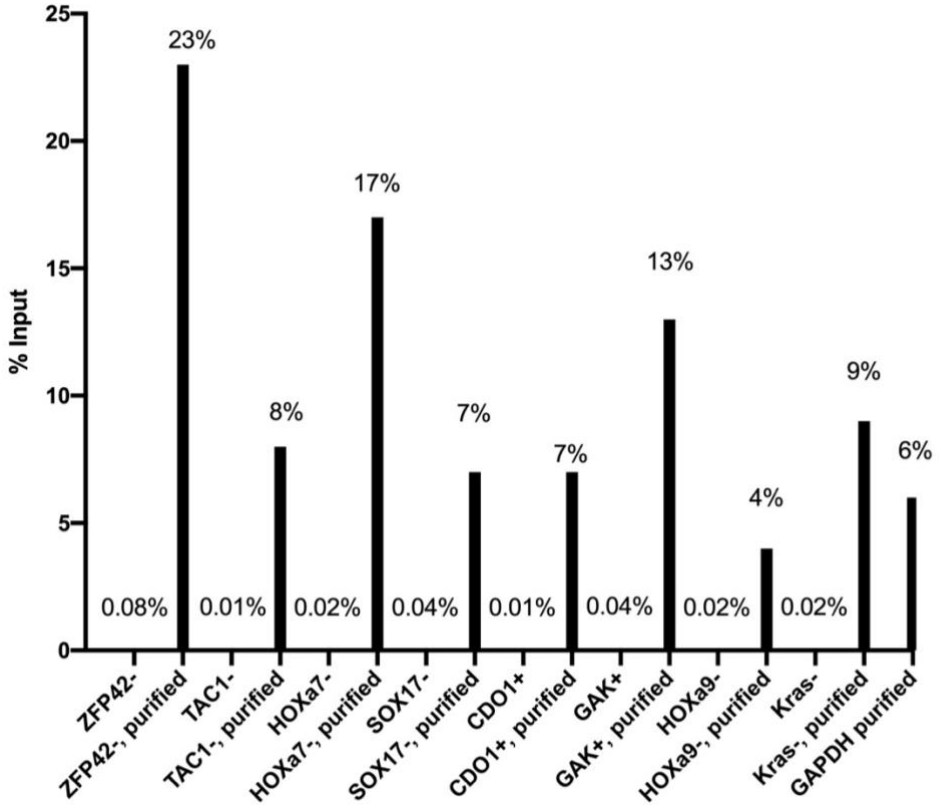

Figure 16 ***Gargle- DNA Purification.*** *The y-axis represents the percentage of captured DNA normalized to the input, while the x-axis depicts the region, along with different conditions for each sample. Performing multiplexing of hybridization with 2.22µL per bead type, comparing unpurified and purified gargle samples. Purifying DNA from gargle samples after digestion has a significant impact on capture efficiency.*

4.1.8 A549: Purified vs. Unpurified

With the gained knowledge from the purification experiments with gargle DNA which showed a much higher capturing ability compared to unpurified gargle, we aimed to repeat the hybridization test with A549 lung cancer DNA to identify if purification of used cancer DNA makes a difference, too. For multiplexing of hybridization, we again mixed 2.22µL from each bead type. Data shown in Figure 17 indicated that the impact of purified DNA is not as high as with the gargle samples, though some percentages get even doubled like for TAC1-, HOXa7-, SOX17-, GAK+. The %input of HOXa9 and KRAS DNA is more than double for the purified conditions. ZFP42- and KRAS- stay the probes with the highest %input results during the testing period.

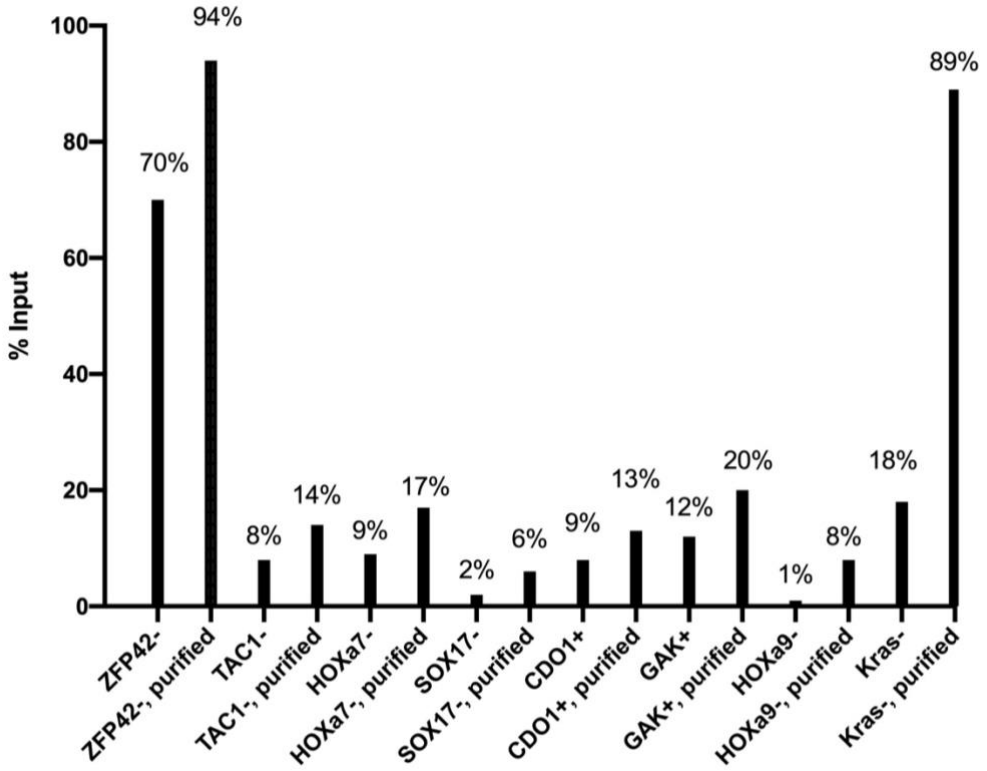

Figure 17 ***A549 – DNA purification****. The y-axis represents the percentage of captured DNA normalized to the input, while the x-axis depicts the region, along with different conditions for each sample... Purified DNA shows better capture efficiency.*

To conclude, the purification of A549 cancer DNA shows a minor difference comparing to the unpurified condition, whereas for gargle it makes a big difference. This can be explained as by purification, we are able to remove of all unwanted substances in gargle samples and so improve the hybridization capture efficiency for the used probes.
In the future, it's crucial to consider gargle purification before hybridization. This significantly improves the capture efficiency, resulting in better target enrichment.

4.2 Validating the functionality of the TAPS protocol with in-house produced mTET enzyme

This chapter demonstrates the effective implementation of the published protocol ([1]), which utilizes the TET enzyme followed by chemical treatment to convert methylated cytosines to uracil, resulting in C→T conversion after PCR. Importantly, this protocol operates efficiently in our experimental context, using an in-house-produced TET enzyme.
In addition to expressing the TET enzyme in insect cells like in the published protocol, we aimed to investigate whether TET enzyme expressed in E. coli would work similarly efficient. Due to the substantially lower cost of E. coli cultures compared to insect cells, expressing the enzyme in E. coli would enable affordable scaling up of enzyme production for the planned high-throughput experiment.
For the following experiments we used genomic DNA (gDNA) extracted from A549 lung cancer cells and used two regions as our controls, the region GAK as a positive control (consistently methylated) and GAPDH as negative control (consistently unmethylated).

4.2.1 Validation of TAPS method with in-house produced TET enzyme expressed in insect cells

To assess the efficiency of the in-house produced TET enzyme (expressed in insect cells), we performed a titration of the concentration ranging from 2µM to 8µM, covering values equivalent to half and double the published concentration of 4µM. In addition to the previously mentioned positive controls GAK and GAPDH, we included the RHBDF region, which, like GAK, displays consistent methylation, and HOXa9 and HOXa11, regions specifically methylated in lung cancer according to the earlier-mentioned publication in this thesis ([2]). The TAPS reaction was performed on 100ng of genomic DNA, and specific regions were subsequently amplified and sequenced using Next-Generation Sequencing (NGS). For the readout we looked at the percentage of TAPS converted CG positions of every region in comparison to the initially methylated DNA sequence.
The experimental outcome is shown in Figure 18. All regions expected to be methylated are converted, and GAPDH, which is not methylated, shows no conversion. Concentrations of 2µM and 3µM exhibit the lowest conversion rates, while concentrations from 4uM upwards demonstrate the best conversions with only minor differences. Consequently, we have decided to use 5µM in all subsequent experiments.

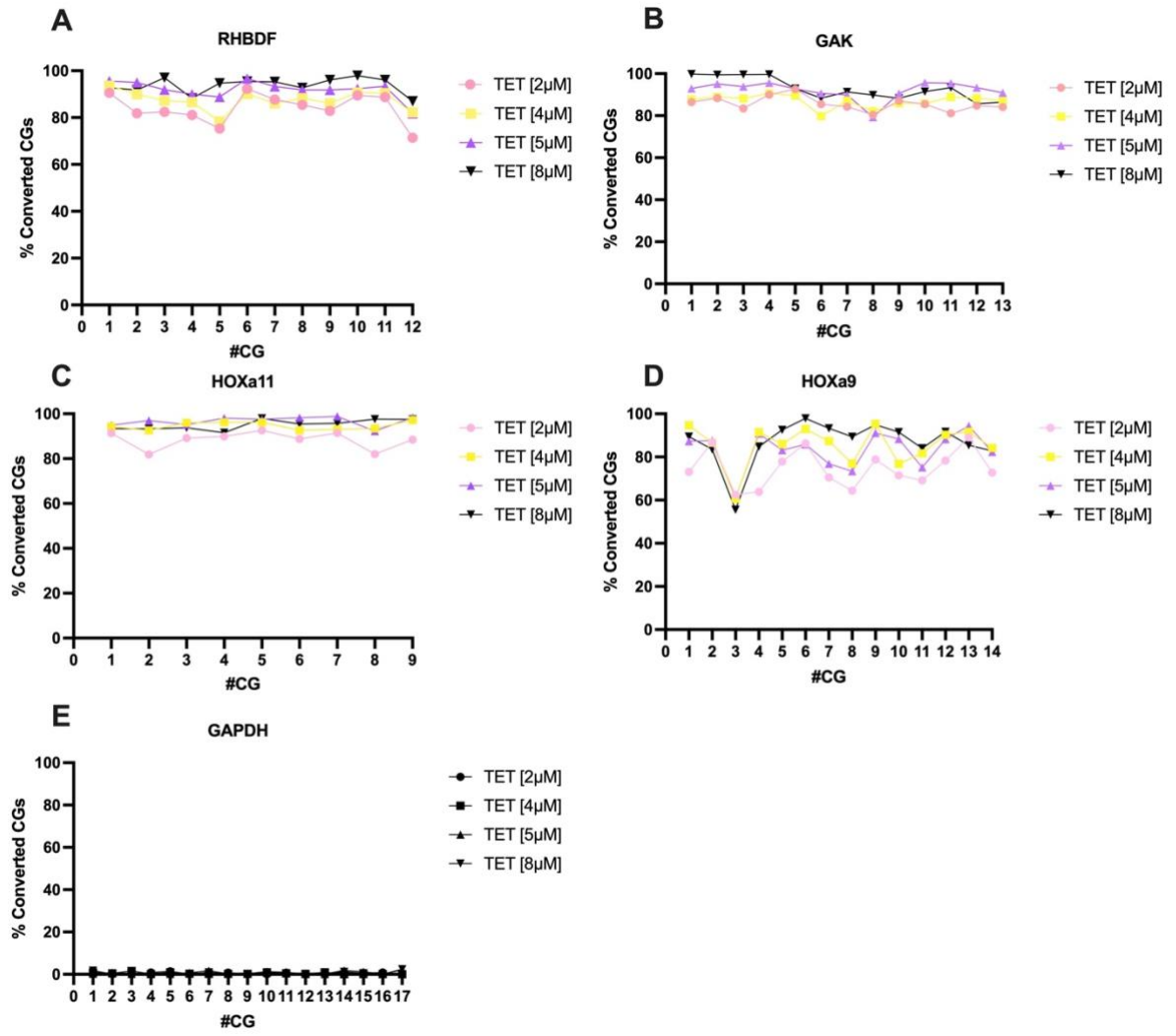

Figure 18 ***TET-efficiency test.*** *A dilution series was set up where different concentrations of enzyme were to test if a higher conversion rate can be obtained with more TET. The results show that there is no big difference even when 8μM TET are used, which would stand for 200% and still does not reach a conversion rate of 100% GAPDH is the negative control and looks as expected with no conversion, at all.*

Using a concentration of 5μM of the TET enzyme, we replicated the experiment for the same regions. This time, the TAPS protocol was performed under two conditions: one sample was treated with TET, and the other was not. This approach aimed to ensure that the conversion occurred specifically due to TET activity rather than being randomly induced by chemical treatment.

GAPDH should exhibit 0% conversion in both cases as it is a non-methylated region, preventing any oxidation reaction. The other loci are expected to display converted CGs in the condition where TET is used, while the opposite should be observed in reactions without TET in the reaction. Figure 19 illustrates the proposed hypothesis, showing 90-95% conversion for the four methylated regions in TET-treated samples and no conversion in reactions without TET. No conversions were observed in both conditions for the negative control GAPDH.

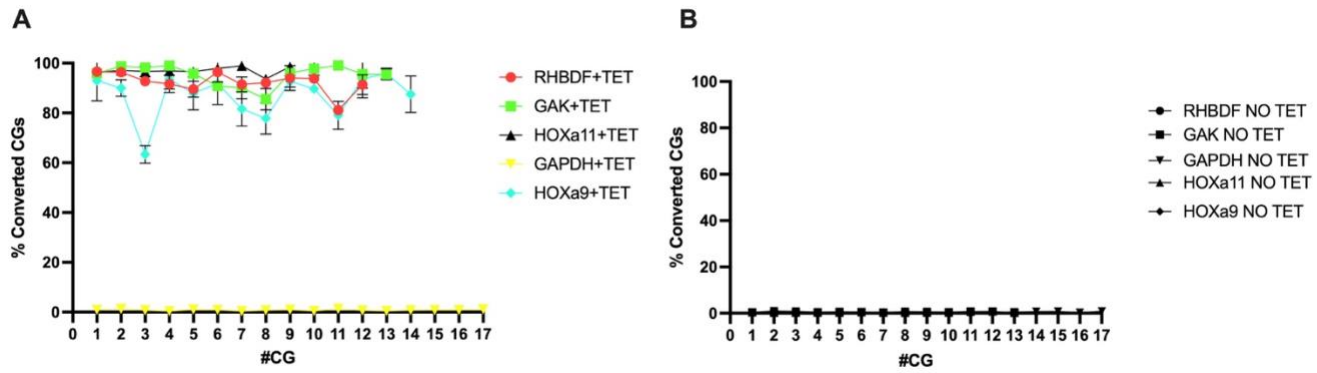

Figure 19 ***Methylation detection.*** *The TAPS method works, it shows an approximate methylation status of 90-95%. The samples treated with TET indicate that overall, 90-95% of CGs were converted during the TAPS protocol, except for GAPDH which has no conversions and serves as our negative control. GAK as the positive control has around 95% converted cytosines in the CG regions. If TET is not spiked in into the oxidation buffer at the beginning of TAPS, there is collectively no conversion to be seen, as it was predicted, therefore no methylated cytosines can get converted because of the lack of oxidation step during the TET reaction. (n=2)*

4.2.2 Validation of TAPS method with in-house produced TET enzyme expressed in E. coli

With the aim of establishing a high-throughput methylation detection method, the only expensive component in the TAPS protocol is the TET enzyme. Therefore, we sought to utilize E. coli for protein expression, given its cost-effectiveness. The in-house protein facility successfully expressed the TET enzyme using E. coli, which we evaluated in a dilution series comparing the efficiency of various enzyme concentrations to the 5µM insect-TET used as a control. In addition to the GAK, GAPDH, and RHBDF1 regions, we included ZFP42 as a lung cancer-specific region this time ([2]).

In Figure 20, it is evident that a lower enzyme amount (2µM) shows almost no conversion for all regions. The higher the concentration, the greater the conversion rate. However, the 4µM concentration of TET enzyme from E. coli does not reach the 95% conversion rate achieved by the control TET enzyme expressed in insect cells. Only when using double the amount (i.e., 8µM) is an even higher conversion rate achieved. This suggests that the TET enzyme expressed in E. coli has slightly less activity, which needs to be compensated for by the higher concentration. The variable rate of conversions for different CG positions, especially in the ZFP42 region, indicates potentially distinct accessibility for CGs by the TET enzyme or different levels of methylation.

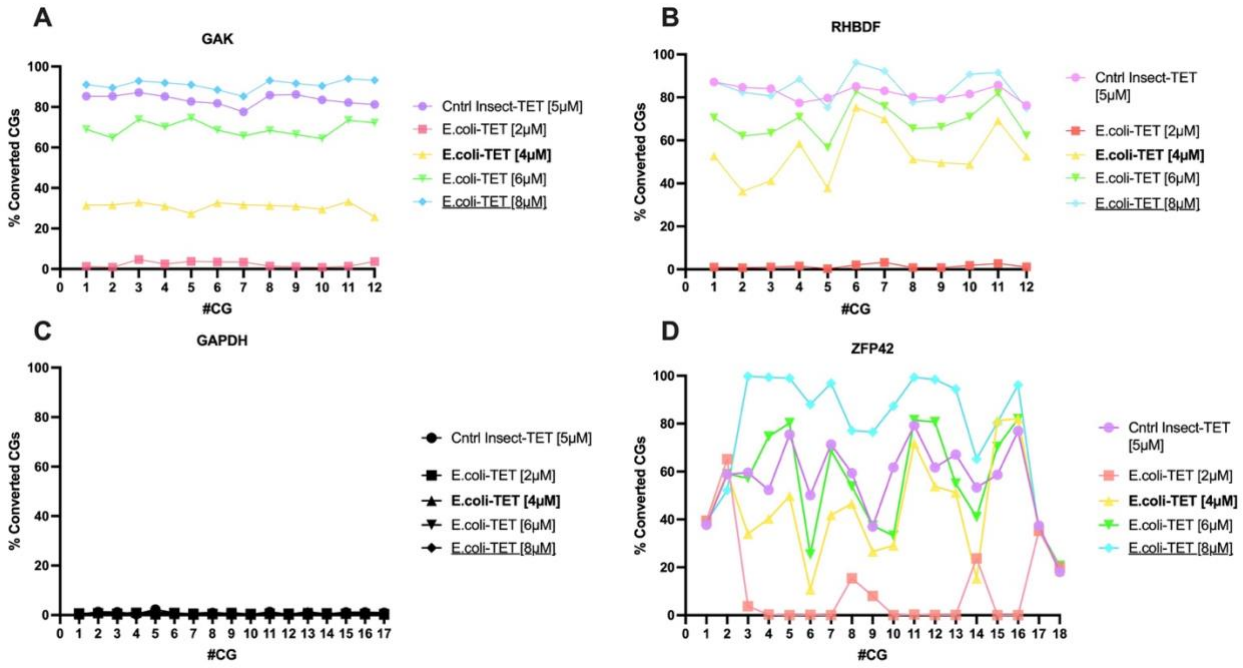

Figure 20 ***TET dilution series.*** *With 8µM TET expressed in E. coli a conversion rate of above 90% is achievable. A) Collectively, one can say the more enzyme is put into the TET oxidation reaction, the higher the conversion rate becomes, with 8µM E. coli TET succeeding in overtaking the 5µM insect-TET and reaching a conversion of nearly 100%. B) For region RHBDF the same can be said as for GAK, mentioning that 8µM E. coli TET reaches a conversion of 100% for some of the CGs. C) GAPDH shows no conversion at all, which indicates the region does not have any methylated cytosines and serves as a good negative control. D) ZFP42 shows to be a more complicated region where CGs are not as accessible as in the other cases, you can see depicted the number of CGs in the given locus where some positions are less converted than others, but in total the use of 8µM TET is once more able to even reach a conversion-status of 100%, indicating that the enzyme is saturated here and perform its function efficiently.*

4.3 TAPS on the beads

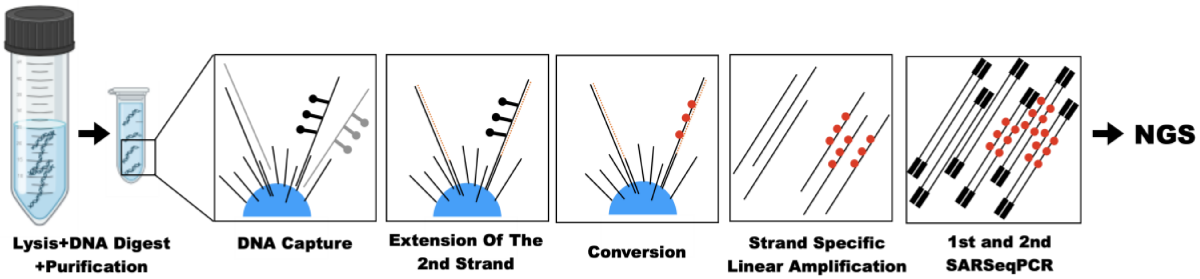

Figure 21 ***TAPS on the beads Workflow****. Depicted you can see the workflow showing firstly the lysis+DNA digest and purification, followed by the target hybridization capture protocol. We build the bridge between hybridization and TAPS by performing a second strand synthesis by extending and filling up the capturing strand in order to create a double stranded DNA for the TET enzyme to be able to work on. After the conversion step via the TAPS protocol, we perform a strand specific linear amplification where we want to amplify only the target strand which gets captured by the beads and is initially methylated, using only one primer (either the one for the FW or the REV strand). The next step is the 1st SARSeq PCR where the linearly amplified single strands get now processed in a PCR where primers for both FW as well as REV are used in order to create double stranded DNA again. The 2nd SARSeq PCR is used to incorporate barcodes into the DNA in order to be able to perform multiplexing of all the samples for later identification distinction of DNA samples during NGS analysis.*

This chapter explains how the TAPS protocol, as outlined in Chapter 3.2.5, was applied to the captured DNA (described in Chapter 3.2.4). Before conducting TAPS on the captured DNA (i.e., TAPS on the beads), a secondary strand extension was necessary to stabilize the DNA on the beads and enhance the efficiency of the TET enzyme as it preferentially acts on dsDNA. The resulting synthesized strand lacks methylations. Consequently, for Next-Generation Sequencing (NGS), it was important to amplify only the original strand (Figure 21). We explored various strategies to selectively obtain the converted strand which are explained in the next paragraph. Additionally, we conducted tests to assess whether the addition of 1% Triton and/or 0.1 mg/ml BSA would enhance the efficiency of the TET enzyme.

All the experiment were performed on A549 DNA.

From this point onward, we have implemented a distinct analysis approach for the NGS data. Instead of analyzing the proportion of converted C or G at each position across all reads, we now calculate the fraction of converted Cs or Gs (depending on whether the plus or minus strand was captured) per individual read. This modification is particularly advantageous for detecting minority methylated DNA within the majority of unmethylated DNA.

4.3.1 Strategies for selective amplification of the converted DNA strand

Two methods were tested to separate the converted, non-covalently bound strand from not-converted, covalently bound strand (which remains with the beads). The non-covalently bound strand can enter the solution and be separated from the bound strand by removing it from the beads, enabling the selective amplification (1st and 2nd SARSeq PCR) of only the converted strand. The 1st SARSeq PCR introduces well barcodes along with i5 and i7 Illumina indexes using forward and reverse primers. This is followed by the 2nd SARSeq PCR, which introduces the plate barcode along with P5 and P7 Illumina adapters (see methods).
The two strategies were:

1. Different elution conditions:

Various elution strategies were explored. Elution was considered effective for our purpose because the non-methylated strand is covalently bound to the beads due to the extension process (see Fig 21).

Instead of utilizing the elution method with NaOH at room temperature (as done for the qPCR readout after hybridization (see methods), alternative conditions were tested. This included elution with NaOH at 60°C and at 95°C. The elevated pH and temperatures were expected to aid in denaturing the two long DNA strands, allowing the converted strand (not covalently bound) to enter the solution for subsequent amplification by the two PCR (1st and 2nd SARSeq PCR).

2. **Strand-specific linear amplification:**

For strand-specific linear amplification, a primer designed to amplify only the converted strand was used. The SARSeq PCR1 using the second primer was conducted on the reaction without the beads, as the linearly amplified strand is expected to be in the solvent. The 2nd PCR was carried out as usual.

Figure 22 illustrates the various strategies used to enable the selective amplification of the converted strand, perfomed with GAK and ZFP42 minus strand capturing beads. The results clearly demonstrate that utilizing the single-strand linear amplification method produces the most effective and strand-specific NGS results. In the case of both GAK and ZFP42, the majority of reads using the single-strand linear amplification method have 10-12 converted Cs per read for GAK and 10-14 converted Cs per read for ZFP42.

For the elution strategy, particularly elution at 95°C, it is evident in both cases that the majority of reads contain no converted Cs, suggesting potential degradation of the converted strand. Elution with NaOH at room temperature (RT) and 60°C reveals the presence of both unconverted and converted reads.

In conclusion, the strand-specific linear amplification yielded the most optimal results, and therefore, we will proceed with this method.

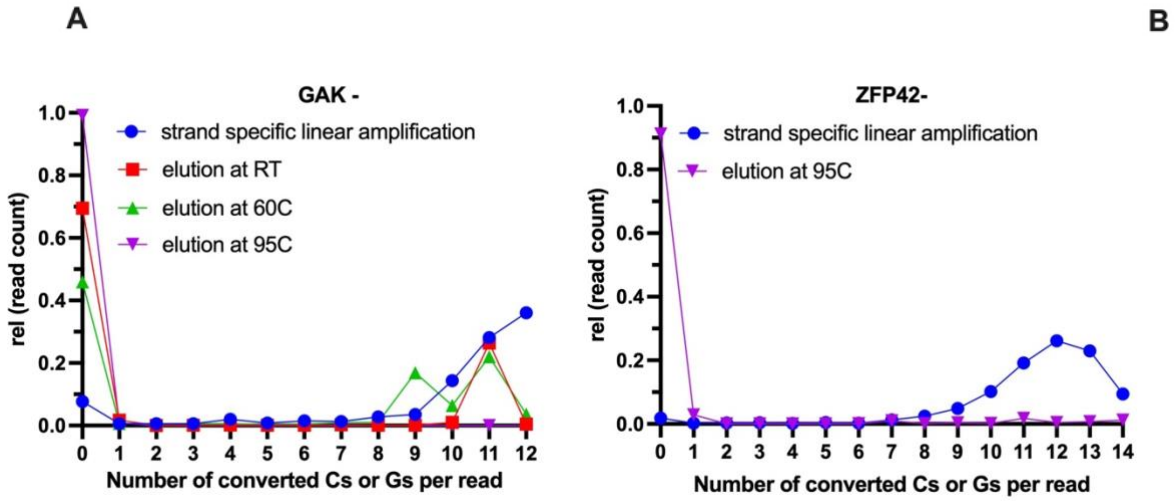

Figure 22 ***Elution methods.*** *Strand specific PCR (ssPCR) followed by a standard PCR show that in the majority of reads the region specific number of methylated CGs gets fully converted. A) For GAK- different elution conditions with NaOH for different incubation temperatures were tested opposite of the ssPCR strategy. The green coloured plot indicates the succesful ssPCR method. B) For ZFP42 only the standard elution with NaOH at room temperature was tested, bringing us the same result as for GAK. The given reads show no conversion for the standard elution, but with the strand specific PCR and following PCR strategy we are able to successfully amplify our captured target strand and getting the information that the reads show a conversion of the methylated CGs in the amplified region (blue coloured plot).*

4.3.2 Testing if Bovine serum albumin (BSA) or Triton increases the TET-efficiency

After identifying the optimal elution strategy with the strand specific linear amplification, we addressed next whether adding 0,1 mg/mL BSA (end concentration) or 1% Triton (end concentration) to the TET oxidation buffer would increase the efficiency of E. coli-TET. Beads capturing the GAK- strand were used for this test. The experiment design included the use of 4µM, 8µM and 12µM TET enzyme per reaction, comparing them to 4µM with 0,1 mg/mL BSA and 4µM with 1%Triton. As a negative control a condition without the TET was chosen. The results in Figure 23 shows that BSA improves the TET efficiency, as the majority of the reads show a conversion of more than 9 Cs per read, with only a minimum number of reads showing no conversion. This pattern closely resembles the conditions with 8 µM and 12 µM TET.
This let us conclude that we were able to save enzyme and use 4µM in combination with BSA to reach the same effect as if twice the amount of TET was spiked in into the reaction.

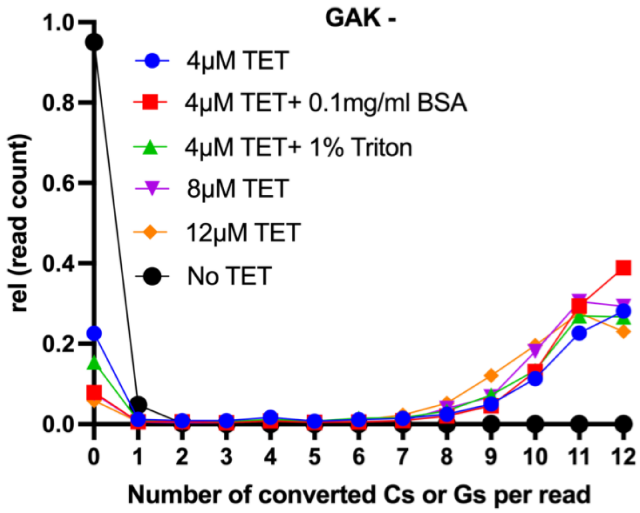

Figure 23 ***TET oxidation optimization.*** *4μM TET in combination with BSA shows same results as the use of 8μM or even 12μM enzyme. The GAK samples show we get most of the reads indicating a successful and full CG conversion when using less E. coli-TET and add BSA into the oxidation buffer. When no TET is used there are only reads with information that there are no converted CGs in this sample to find.*

Not only is BSA making the condition for the enzyme better, but Triton also helps as well with the washing of the beads between the different stages of TAPS by preventing the clumping of the beads on the walls of the tubes and aiding them to get in solution more smoothly.

4.3.3 TAPS on the bridge-capturing beads

The TAPS on the beads protocol was optimized in terms of introducing the ideal elution strategy, enhancing the TET enzyme efficiency by developing a better environment for the oxidation buffer and improving the washing steps by avoiding the clumping of beads in the reaction tube between various reaction stages. Following the protocol optimization, which so far was only done on linear-capturing beads, the next step was to test the bridge-capturing beads in the TAPS reaction. The experiments included TAPS reaction on biological replicates of the ZFP42 region, with one replicate for the condition without TET and four biological replicates for conditions with TET.The results are indicated in Figure 24, where samples treated with TET show that the majority of the reads have between 10-14 converted Cs per read. In contrast, the condition without the TET enzyme reveals that the majority of the reads have no conversions.

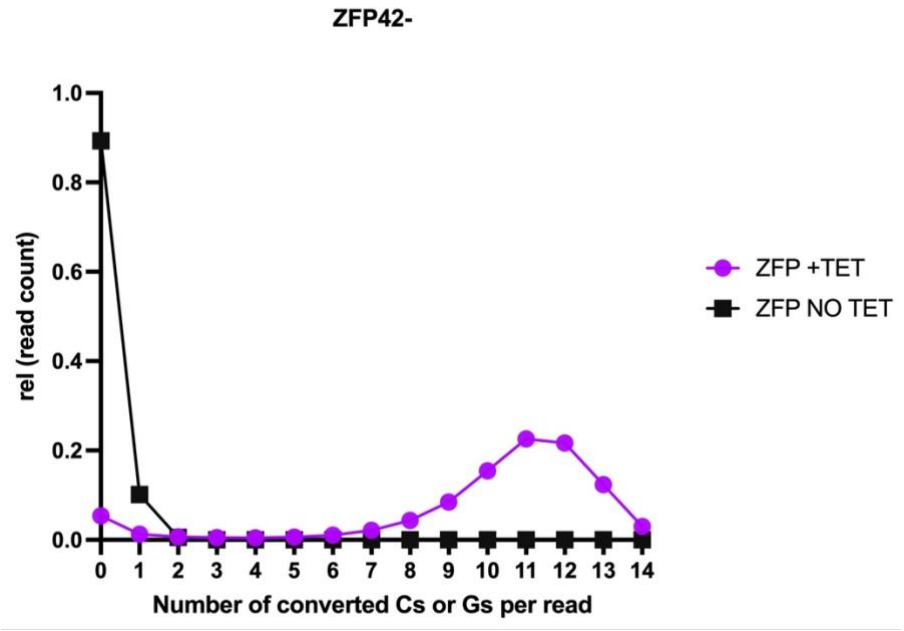

Figure 24 ***ZFP42-Bridge-capturing with TAPS on the beads.*** *The per read analysis shows that we get the most reads with 10 to 14 converted Cs pre read and indicates a successful conversion of the methylated CGs for region ZFP42. In contrast, the reads for 'NO TET" treatment are the highest for no converted CGs.*

In the next experiment we included the beads for the HOXa9 region and tested both conditions, one with TET and one without TET.

The outcome for both conditions is shown in Figure 25 where we were able to reproduce the same results for HOXa9 as for ZFP. In both cases the samples without TET treatment show that none of the methylated CGs are converted due to the lack of the oxidation activity by the enzyme, whereas the '+TET' treated DNA is converted in 18 of the 21 CGs.

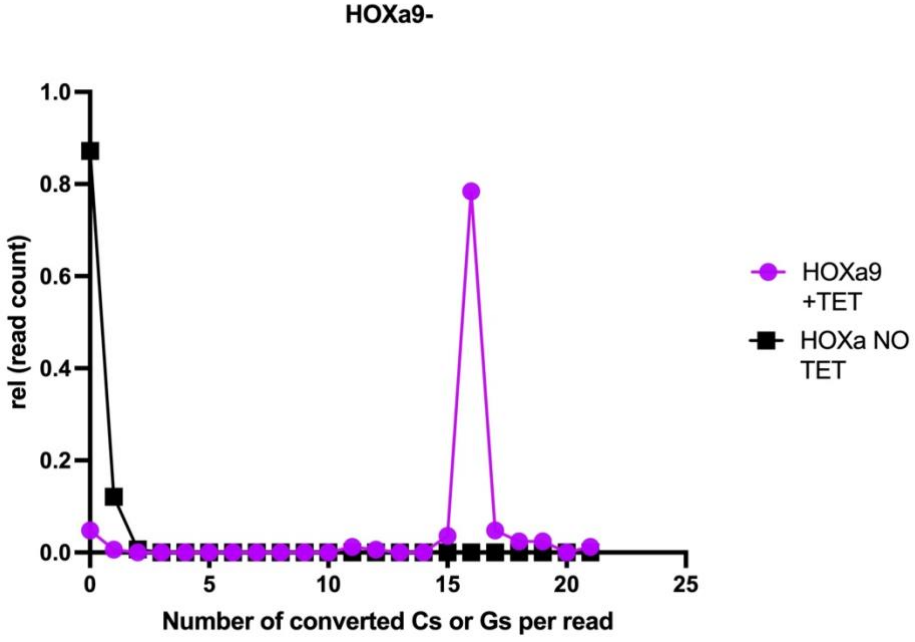

Figure 25 ***HOXa9-Bridge-capturing with TAPS on the beads.*** *Testing another region in HOXa9 mirrors the great results from the ZFP plot and the high read count indicates that the 18 of the 21 CGs are methylated and converted after TAPS. The "HOXa9 NO TET" negative control mirrors the same condition as in Figure 26, with no TET oxidation taking place and therefore the CG status is kept throughout.*

In conclusion, TAPS on the bridge-capturing beads works as effectively as on the single-capturing beads.

4.3.4 Multiplexed TAPS of beads

The “TAPS on the beads” protocol has been optimized and is working efficiently on both linear and bridge-capturing beads. Next step involved performing TAPS on multiplexed bridge-capturing beads. TAPS is applied for the first time for the lung cancer specific loci ZFP42, HOXa7, HOXA9, GAK, SOX17, TAC1, Kras, as well as GAPDH, being a negative control. Hybridization and TAPS until the borane reduction reaction was done in a multiplex manner. Subsequently, strand-specific linear amplification, as well as the 1st SARSeq PCR, were conducted separately for individual loci and then submitted for Sanger sequencing. Analyzing the sequences brought the conclusion that the various regions show a different methylation pattern indicating that the CG methylation frequency differs between the specific genes. In addition to the sequence alignments, the UCSC genome browser was taken as a tool to check for the methylation status of the sample compared to the original sequence, in some cases with confirmation of methylated CGs for A549 lung cancer DNA. Figure 26 serves as an example for the ZFP42 region. Shown is the ZFP42 control sequence where no cytosines were converted compared to the tested TAPS converted sample, even showing positions that are already confirmed to be found with the mutation for A549 DNA and saved in the database. A color scheme code gave the information that red is standing for 100% of sequenced molecules are methylated, yellow equals 50% of molecules which are sequenced are methylated, and green would be 0%. The methylation status for the database is represented using a color gradient for 11 different shades. On the left side the methylation probability for stated promoter region for different cell and tissue types.

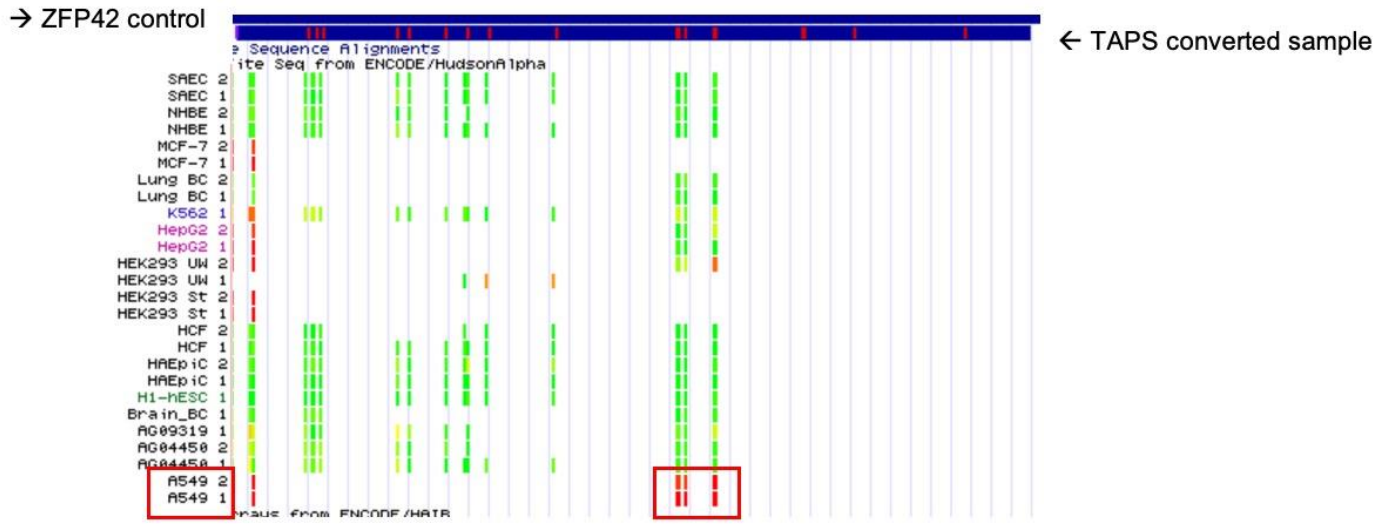

Figure 26 ***ZFP42 - View of UCSC Genome Browser.*** *The red rectangle highlights the conversion of methylated CGs based on DNA Methylation by Reduced Representation Bisulfite Seq from ENCODE/HudsonAlpha in the A549 cancer cell line, which served as the reference dataset for our experiments (REF). ZFP42 control indicates a sample that was not treated with TAPS, while the TAPS-converted sample displays all the converted Cs marked by red lines.*

The sequence alignment pictured in Figure 27 additionally showcases the mutated CGs where a cytosine next to a guanine is converted to thymine, resulting in a transformation from CG to TG, with an excerpt of the peaks received from the Sanger sequencing analysis.

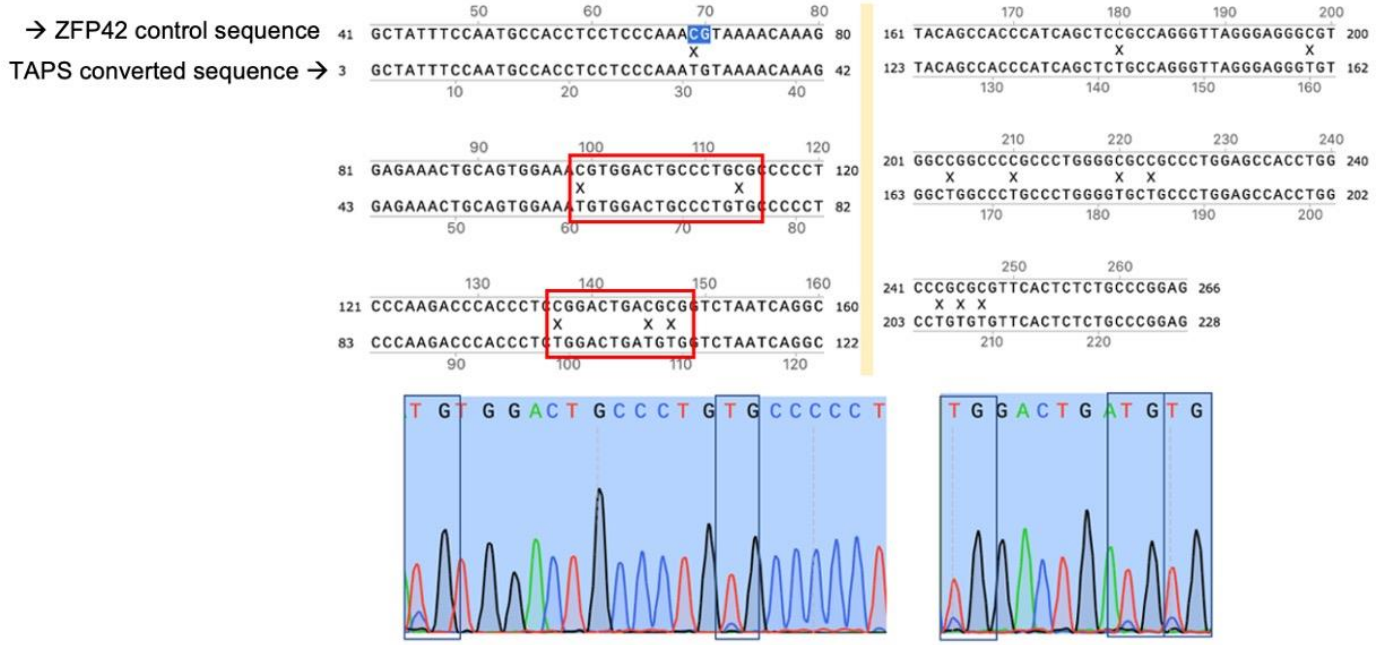

Figure 27 ***ZFP42 Sanger sequence alignments.*** *Sequence alignments show the successful conversion from CG to TG for every position in the ZFP42 region. An excerpt of the sequence from the Sanger analysis below is depicted for the specific TG peaks. (Alignments done with SnapGene)*

For region HOXa7 the databank alignment is to be found in Figure 28, and Figure 29 leads to the conclusion that not every CG position is converted because the methylation pattern is

different than in the highly converted locus of ZFP42 where all CGs are methylated according to the results. With HOXa7 we see a more unique picture of its methylation level. USCS genome browser is also indicating that there is no data of 100% methylated and sequenced molecules, the color code here is more orange to darker yellow in comparison to ZFP for example.

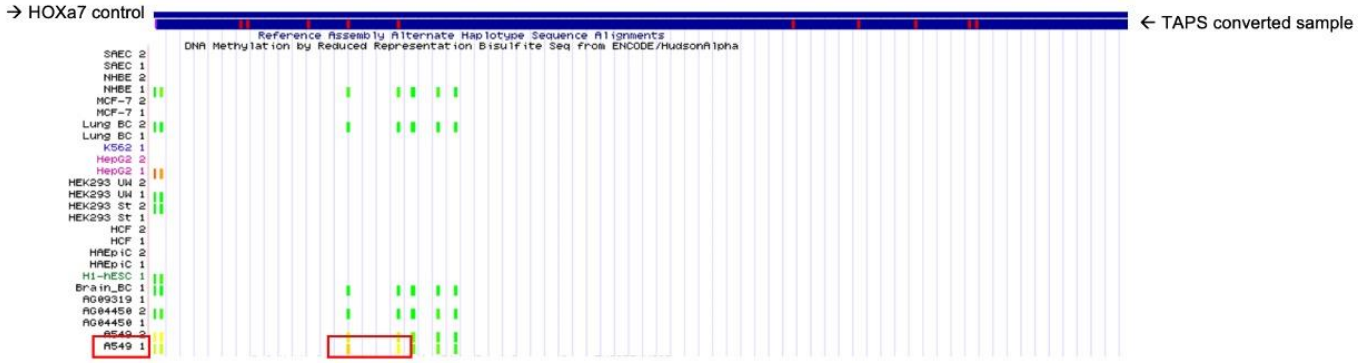

Figure 28 ***HOXa7 - View of UCSC Genome Browser .*** *The TAPS converted sample shows the methylation status compared to the control sequence. For A549 DNA the databank has information about some methylated CGs but the color shade tells us that these mutations are not found in 100% of the sequenced molecules of which the data is used for the UCSC Genome Browser.*

Figure 29 ***HOXa7 Sanger sequence alignment****. There are CGs converted to TGs in many cases but there are also positions which are not, indicating that not every cytosine next to a guanine is probably methylated. (Alignments done with SnapGene)*

Figure 30 supports the results from the NGS analysis in Figure 25 where it was shown that for HOXa9 18 of the 21 CGs in the amplified region are converted, proofing once more that these are regions with a specific methylation pattern and not every single cytosine in the DNA section is mutated. The Sanger sequences alignment in Figure 30 depicts the sequence with 3 CG positions that stayed CG and did not become a TG after TAPS (highlighted in blue in the figure).

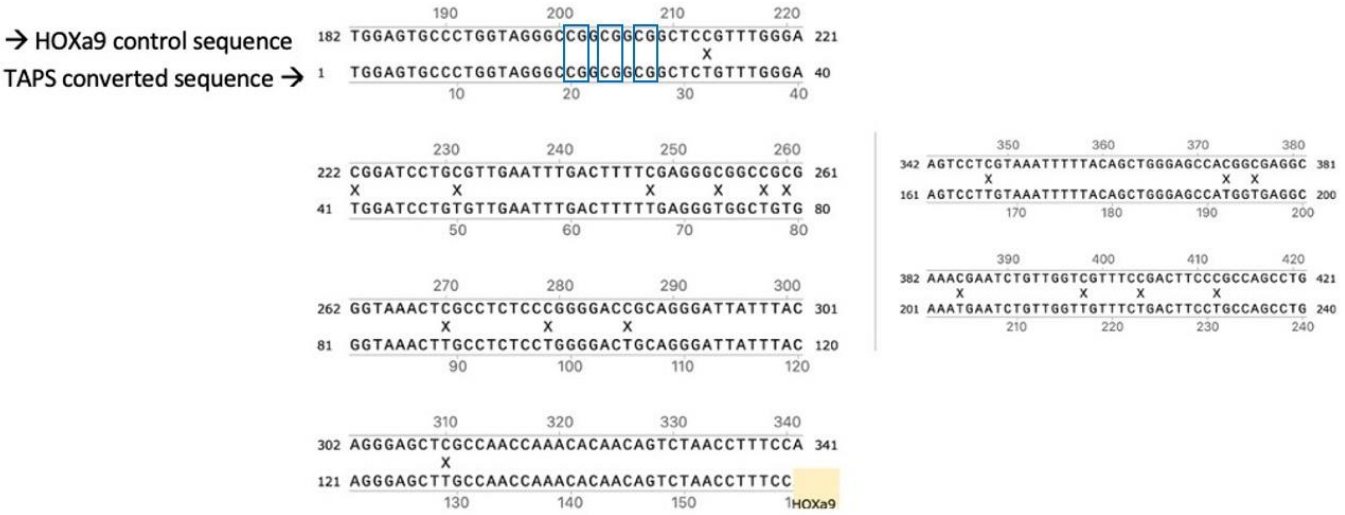

Figure 30 ***HOXa9 Sanger sequence alignment.*** *A seen 18 of 21 CG positions show a TAPS conversion from CG to TG. (Alignments done with SnapGene)*

The GAK region serves as a positive control. In view of the fact that for GAK we use a bridge-capturing bead which is designed for capturing the +strand, we will see a conversion from CG to CA in this case. This and the full conversion of every cytosine in a CG position after TAPS is well and clearly depicted in Figure 31. An excerpt of the peaks generated by the Sanger sequencing analysis are showcasing its successful and complete CG-conversion into CA (Figure 32).

Figure 31 ***GAK - View of UCSC Genome Browser .*** *The positive control proofs the successful TAPS procedure as every CG is well known to be methylated in this region. Th database is even confirming the methylation status for A549 lung cancer DNA.*

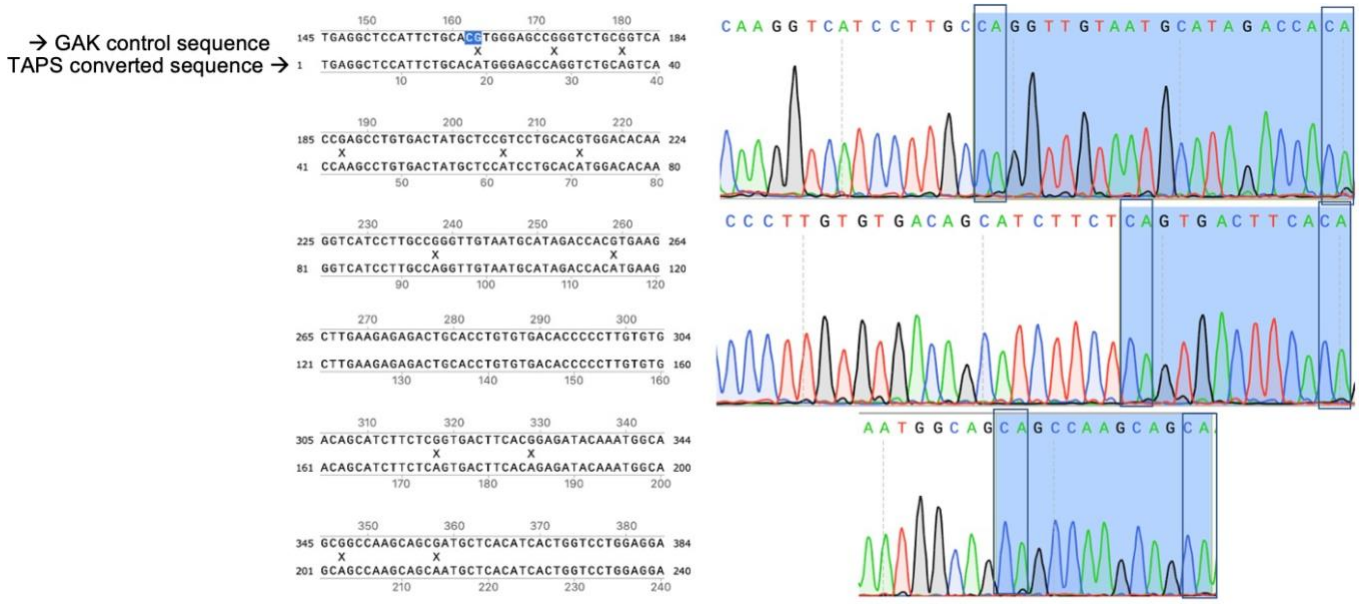

Figure 32 ***GAK Sanger sequence alignment.*** *Very nicely depicted the TAPS conversion from CG to CA, for every single CG position. A successful TAPS performance confirmed by this positive control. (Alignments done with SnapGene)*

The next region which is SOX17 is stated as a lung cancer specific gene, in our experiment this locus showed no conversion after TAPS, as seen in Figure 33 where the database is also not indicating sequenced molecules that were methylated in stated promoter region of A549 lung cancer DNA. Figure 34 presents the alignment after Sanger sequencing analysis.

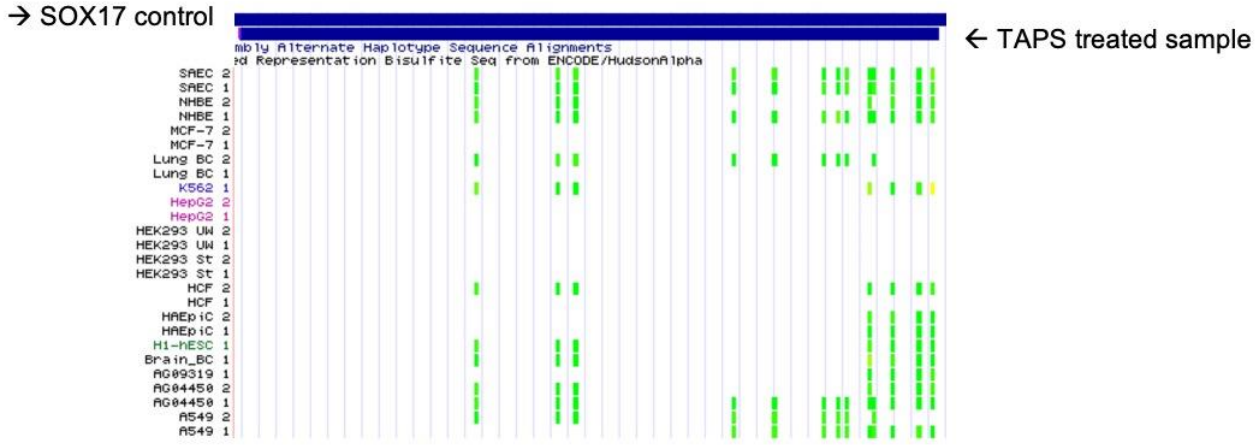

Figure 33 ***SOX17 - View of UCSC Genome Browser.*** *As described, no TAPS conversion is to be seen for SOX17, even indicating no sequenced molecules with a methylation pattern for A549 lung cancer by database.*

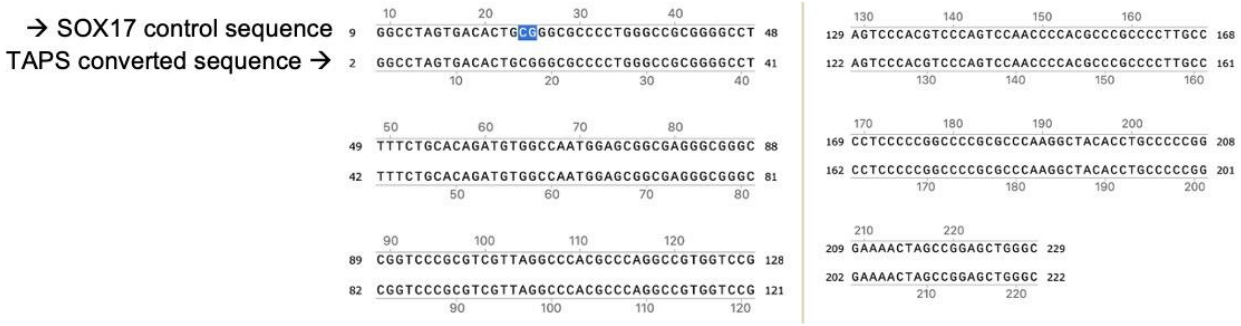

Figure 34 ***SOX17 Sanger sequence alignment.*** *No CG conversion to be seen after TAPS.*

At first, the aligned data for region TAC1 showed no conversion in the UCSC Genome Browser, but the deeper analysis of the Sanger sequence showed clearly double peaks for the CG positions, indicating that the software still identified them as CGs but a T in every C(G) is appearing as one can recognize in a sequence excerpt in Figure 35.

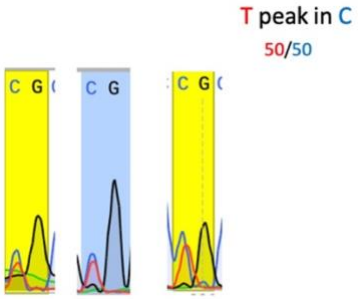

Figure 35 ***TAC1 Sanger sequence peaks.*** *50/50 peaks after Sanger sequencing for TAC1, indicating that TAPS did work.*

Kras is a very short region of around 100 bp. It was chosen as a region because of its mutation in A549 lung cancer, where additionally to the CG methylation another G12S mutation is observed where guanine is exchanged by adenine, see Figure 37. So, not only do we have TAPS converted cytosines, but we as an additional data also have the A549 specific Kras mutation, too. At first sight, the databank of the UCSC Genome Browser is not telling that there are methylated Cs which are TAPS converted with a red mark, as can be observed in Figure 36, although a difference to the original sequence with an indicated 'T' can be spotted, slightly downstream the G12S mutation, which can be linked to the alignment in Figure 37. Further analysis of the sequences received from Sanger detects converted CGs to TGs which can be taken from the peaks and alignment in Figure 37.

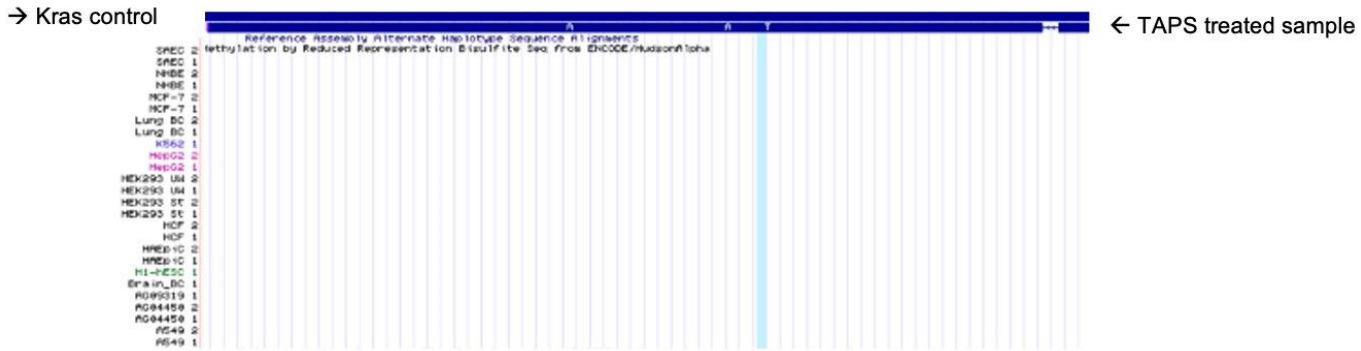

Figure 36 ***Kras - View of UCSC Genome Browser .*** *At first sight, no conversions can be detected from the database, therefore we are taking a closer look at the sequences in Figure 41.*

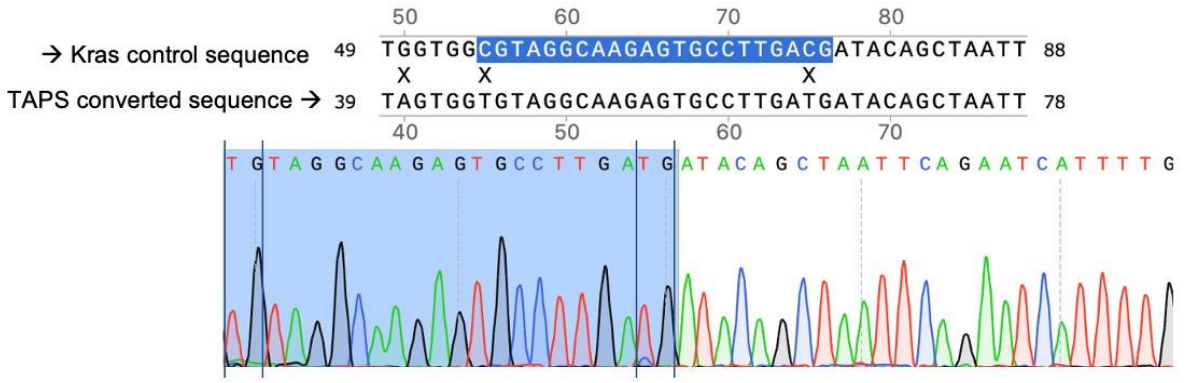

Figure 37 ***Kras Sanger sequence alignment.*** *Sanger sequencing peaks for the Kras TAPS conversions, CG does become TG, indicating the methylation status of the region for A549.*

Region GAPDH did not show any CG conversion, proofing the negative methylation status of the locus and praising it to be an excellent negative control. An excerpt of the analyzed sequence after Sanger is shown in Figure 38.

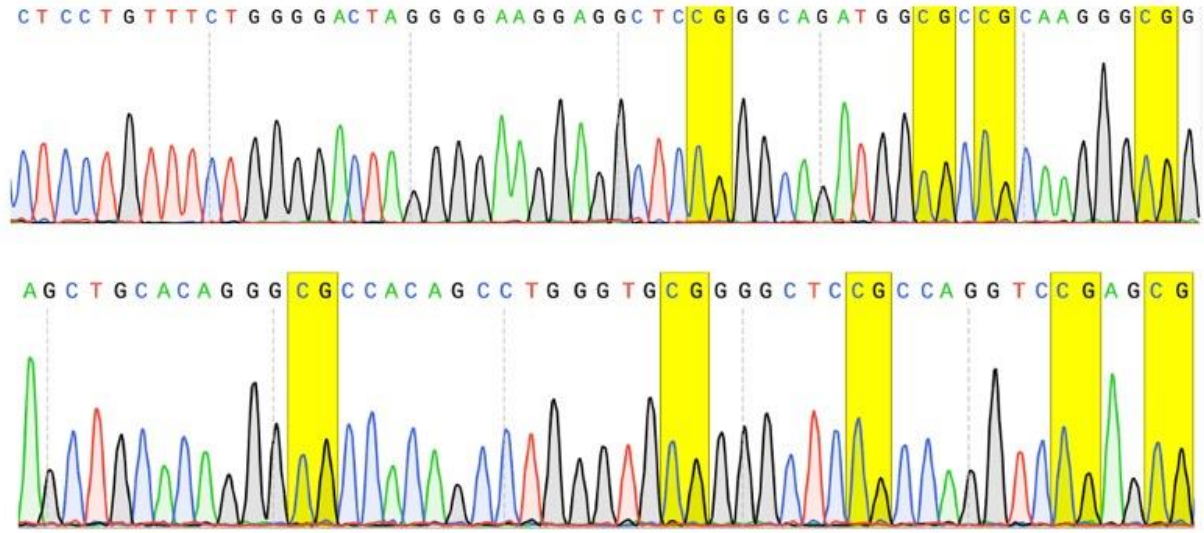

Figure 38 ***GAPDH Sanger sequence.*** *GAPDH sequence after TAPS analysed by Sanger Seq., confirming its high quality as a negative control for this experiment. Every CG stayed a CG because of the lack of methylation.*

In conclusion, it is to mention that the experiment showed a successful outcome in terms of introducing a combined multiplexed hybridization and TAPS reaction. The subsequent PCR for each locus depicted the difference between the various lung cancer specific regions of interest concerning their methylation frequency, as every single locus has its individual methylation status. ZFP42, for instance, is highly methylated, as well as HOXa9 but for HOXa9 there are two accomplished experiments showcasing that 3 of the 21 CG positions are not converted after TAPS, because of their assumed lack of methylation.

4.4 Oral Keratinocytes

This chapter will focus on the task of trying to find an alternative non-methylated genomic DNA to gargle in order to perform and plan a future sensitivity test, where various amounts of cancer DNA is spiked in. Having oral keratinocytes from mouth epithelium in our lab, very suitable as they also could find their way into a gargle sample during its sampling, we choose them to confirm the assumption they do not have methylated CGs in our regions of interest and set up a TAPS in solution experiment to proof our theory. The same loci are tested for their methylation status as in the last multiplex experiment in 4.3.4. GAK serving as a positive control was expected to show a full conversion after TAPS due to the fact that it is methylated in oral keratinocytes, too, and there we should see a conversion of every single CG found in this region, which exactly got confirmed and can be observed in Figure 40. All the other loci including the negative control GAPDH were sequenced by NGS and collectively for all read counts showed that we are working with non-methylated DNA in oral keratinocytes (Figure 39).

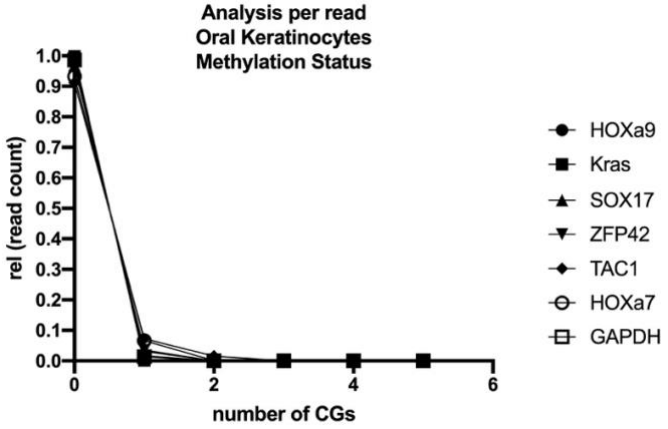

Figure 39 ***Oral keratinocytes methylation status.*** *NGS analysis of lung cancer specific loci in oral keratinocytes, confirming collectively that none of the depicted regions have methylated CGs.*

Figure 40 ***Alignment of GAK after TAPS for oral keratinocyte gDNA.*** *The positive control shows its high level of methylation, in converting all its CG positions during TAPS, some of them being indicated as cancer specific by the BLAT databank in the case of A549, some being only described as methylations which are normally present in healthy cell, like the tested oral keratinocytes.*

5. Discussion

Hybridization

Initially, the long ZFP42 beads designed to capture the minus strand were our starting point. Through protocol optimization, we learned that with the use of a rotator at room temperature as our hybridization condition, a single wash with 70%-EtOH and heat elution on the magnet proved to be the most effective for our protocol. Additionally, we found that a pre-denaturation step involving DMSO and Betaine had a significant impact on the hybridization results. It is highly beneficial in achieving a higher percentage of DNA hybridized to the probes, in the end. A new bridge-capturing type of bead was introduced and showed a higher capturing percentage of DNA. For multiplexing of hybridization, we observed using 2.22µL per beads instead of the standard 10µL for the single-bead reaction resulted in capturing fewer molecules, as expected with a smaller volume of beads, nevertheless, the multiplexing was successful. Importantly, we tested the effect of DNA purification before implementing the hybridization capturing protocol. With the A549 lung cancer cell line gDNA, purification made a noticeable difference. However, the impact was even more pronounced with gargle samples, as potential traces of food, coffee etc. could potentially harm or even inhibit parts of the reaction. In our hybridization process, successfully mixing all the beads together and performing the capture protocol in a single reaction was a significant achievement. This success has brought us one step closer to our goal of combining it with the TAPS method.

TAPS in solution

The TAPS experiments have commenced in solution where we were able to consistently demonstrate the positive conversion effect on various genes. In contrast, the sequencing of samples where the TET enzyme was not spiked in into the oxidation reaction buffer, collectively showed no conversion. The positive control GAK and the negative control GAPDH brought the expected results every time, the first stated region being highly methylated, not only on A549 cancer DNA, the latter having no methylated cytosines in its site, at all. When we carried out the dilution series experiments using TET enzyme which was generated once from insects and once from E. coli organisms we observed that it worked as expected.

TAPS on the beads

Due to the fact that the TET enzyme is only able to work on single-stranded DNA, we needed to introduce a method to generate a second strand synthesis on the beads. We achieved this by introducing an extension protocol after the hybridization and before TAPS.

Furthermore, by optimizing the TET oxidation buffer for the TAPS on the beads protocol with BSA, we achieved the same conversion rate with a standard 4µM concentration as we did with twice the amount of 8µM or even 12µM. The result of this optimization would conclude into significant economic and material savings, as less enzyme is being needed for the experiments. In contrast, the now covalently bound nucleotides made it not possible to get the target strand off the beads. Even various elution attempts using NaOH and heat did not lead to a stable elution method for our purposes. Two subsequent PCRs were able to amplify our strands of interest. First, using a strand specific PCR targeting the captured strand of interest after the borane reaction during the TAPS on the beads protocol followed by a PCR where both strands were again amplified provided us with accurate conversion results. The TAPS on the beads experiments did reveal that the TET enzyme functions on hemi-methylated DNA, too, which we did not know before.

In general, we have succeeded in combining the multiplexing of hybridization where all our beads for the targeted regions are put together in one reaction, lowering the amount of materials as well as the costs at the same time, with continuous TAPS on the beads, as demonstrated by the results from the chapter where we performed a multiplexed TAPS protocol targeted at lung cancer-specific regions of interest. Multiplexing was the goal because we are not working with a lot of material, in general, and having many markers captured in one place already increases the strength of the results. The PCR reactions and readouts were still done separately. In our research, we have further found out that lung cancer-specific genes exhibit a different methylation pattern and frequency for each locus. The effort to find an alternative gDNA without any methylated CGs as a simulation for gargle, which we plan to use for looking further into the sensitivity of our introduced and optimized method, brought us to the successfully analyzed oral keratinocytes from mouth regions. After performing the TAPS protocol on them, none of the target regions in stated keratinocytes showed any conversion, except for GAK which was expected to do so because it served as a positive control, again being highly methylated throughout in this CG rich region.

Oral keratinocytes

After optimizing our protocol and introducing the multiplexing of hybridization as well as TAPS we tested a gDNA which should not have any methylated CGs in our genes of interest. The oral keratinocytes showed no conversion for the lung-specific loci, GAK having its methylation status and serving as a positive control, as expected.

6. Conclusion and outlook

In summary, we have successfully optimized our protocols and introduced a method that could serve as the basis for a future sensitivity test. This experiment was planned because we wanted to test our optimized protocol and study how well it could work. It would be build like this: We would extract, purify and capture two different types of lung cancer DNA. The test would be set up to have a series of different percentages of spiked in cancer into a solution of oral keratinocytes for a first simulation instead of using gargle samples. The chosen series would be laid out for 100%, 50%, 10%, 1%, 0.1% and 0% cancer DNA. In parallel, an in-solution experiment would be prepared, too, to test whether the enrichment of the target regions with our beads are more effective than having all different loci in a solution. Here, the dilution series is set up as followed: 100%, 1%, 0.1% and 0% cancer DNA, only using one type of lung cancer DNA for this purpose. All samples are prepared and treated in triplicates. The PCR reactions for every locus would still be performed separately, with the usual two-PCR system, as already described in the chapters above. The aim of the sensitivity test is that 0.1% of cancer DNA can be detected.

Our next steps involve optimizing the multiplexing of PCR where all primers for all loci are combined in a pool and only one PCR reaction is performed per initial sample, aiming at having one reaction per patient after the development of the finalized method is successful. After the completion of the sensitivity test, the next steps would be to study how much lung DNA we can truly obtain from gargle, because of the presence of other DNA types as mouth, tongue etc. Finally, the ultimate goal is the performance of a sensitivity test using real gargle samples from lung cancer patients and comparing them to the samples from healthy participants. This will allow us to have a deeper look into the method's sensitivity when working with samples from real patients at different stages of the disease.

7. List of figures

8. List of tables

9. References

1. Liu, Y. *et al.* Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* 37, 424–429 (2019).

2. Hulbert, A. *et al.* Early Detection of Lung Cancer Using DNA Promoter Hypermethylation in Plasma and Sputum. *Clin. Cancer Res.* 23, 1998–2005 (2017).

3. Oreskovic, A. & Lutz, B. R. Ultrasensitive hybridization capture: Reliable detection of <1 copy/mL short cell-free DNA from large-volume urine samples. *PLoS ONE* 16, e0247851 (2021).

4. Özkan, E. *et al.* High-throughput Mutational Surveillance of the SARS-CoV-2 Spike Gene. *medRxiv* 2021.07.22.21259587 (2021) doi:10.1101/2021.07.22.21259587.