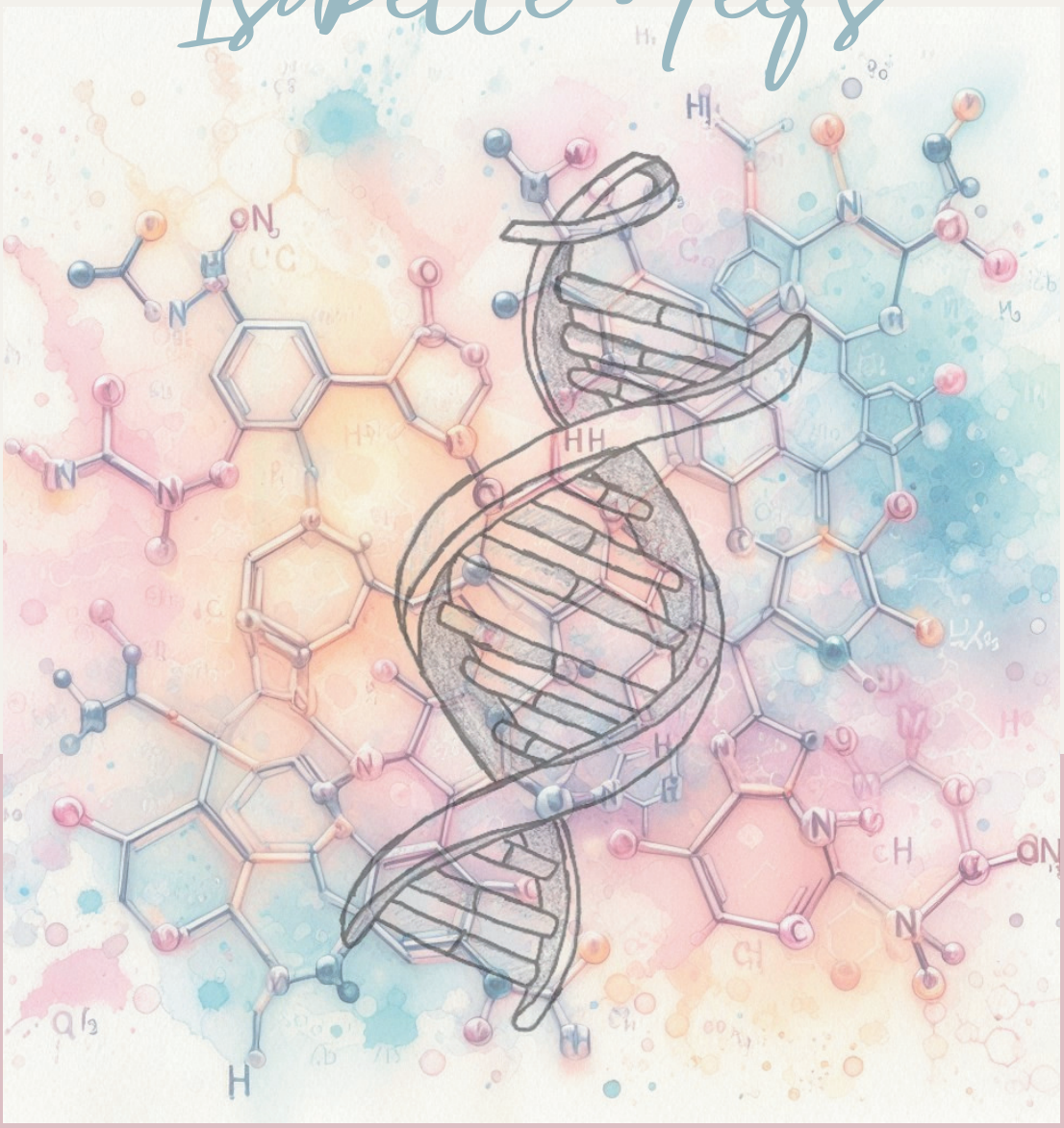
Isabelle Neefs



Enhancing cancer detection through novel DNA methylation strategies and biomarkers

Isabelle Neefs

Enhancing cancer detection through novel DNA methylation strategies and biomarkers

The research was financially supported by the Research Foundation – Flanders (FWO).

Artwork cover by Christophe Neefs & Bing Image Creator
Layout cover by Natacha Hoevenaegel

Faculteit Farmaceutische, Biomedische en Diergeneeskundige Wetenschappen

Departement Biomedische Wetenschappen

Enhancing cancer detection through novel DNA methylation strategies and biomarkers

---

Het verbeteren van kanker detectie via nieuwe DNA methylatie strategieën en biomerkers

Proefschrift voorgelegd tot het behalen van de graad van

Doctor in de Biomedische Wetenschappen

aan de Universiteit Antwerpen

Te verdedigen door

Isabelle Neefs

Promotoren:

Prof. Dr. Ken Op de Beeck

Prof. Dr. Guy Van Camp

Prof. Dr. Marc Peeters

Begeleider:

Prof Dr. Wiebren Tjalma

Antwerpen, 2024

MEMBERS OF THE JURY

Promoters

Prof. Dr. Ken Op de Beeck *(University of Antwerp & Antwerp University Hospital)*

Prof. Dr. Guy Van Camp *(University of Antwerp & Antwerp University Hospital)*

Prof. Dr. Marc Peeters *(University of Antwerp & Antwerp University Hospital)*

Mentor

Prof. Dr. Wiebren Tjalma *(University of Antwerp & Antwerp University Hospital)*

Internal Doctoral Committee

Prof. Dr. Wim Vanden Berghe *(University of Antwerp)*

Prof. Dr. Christophe Deben *(University of Antwerp & Antwerp University Hospital)*

External Jury

Prof. Dr. Katleen De Preter (*University of Ghent*)

Dr. Charlotte D'Hulst (*IMEC, Leuven*)

Table of Contents

TABLE OF CONTENTS

Summary | Samenvatting

SUMMARY

Cancer continues to be a significant global health concern, despite notable advancements in biomedical research. Where most cancers are easily treatable when detected early, finding correct therapies can become challenging in later stages. Timely screening plays thus a pivotal role in improving outcomes for cancer patients. However, traditional diagnostic methods such as physical exams, imaging, and tissue biopsies are often invasive, expensive, and not always precise. Although various innovative diagnostic techniques have emerged, the quest for both reliable biomarkers and sensitive technologies remains ongoing. Among the biomarkers, DNA methylation has emerged as a promising epigenetic factor with significant impact in cancer biomarker research. Hence, investigation into DNA methylation as a potential tool for early cancer detection is crucial.

In response to growing interest, considerable advancements have been made in detecting DNA methylation, particularly in the context of liquid biopsies and cell-free DNA (cfDNA). Traditional methods focused on genomic DNA, but recent innovations have shifted towards enzymatic methods, especially for cfDNA analysis. This has led to a plethora of both classical and novel DNA methylation detection methods, including clinical applications certified by the European *In Vitro* Diagnostics regulations (CE-IVD). Additionally, future trends emerge, such as the potential transition to enzymatic conversions as the new standard and the integration of direct sequencing methods like ONT-sequencing into epigenetic research. Also, the integration of multi-omics technologies becomes more important, which hold promise for enhanced clinical applications by combining diverse biomarkers for improved diagnostic accuracy.

In the second part of this PhD thesis, we focused on novel biomarker discovery, particularly in colorectal cancer (CRC). In a first study, we dived deeper into the CRC screening program and its problems. In particular, we looked at interval cancers (ICs), as they pose a significant challenge in CRC screening initiatives. Previous investigations by our collaborators already shed light on ICs detected by the fecal immunochemical test (FIT), highlighting their prevalence in specific demographic and tumor characteristic subsets. Building upon these insights, our study delved into stage IV CRC patients to elucidate factors associated with FIT-IC. Through analysis of clinicopathological and molecular characteristics of 500 CRCs, we identified associations between the clinicopathological and molecular characteristics and FIT-IC stage IV CRC. Specifically, FIT-IC CRCs exhibited significantly higher odds of being neuroendocrine tumors (NETs) (OR= 5) and displaying lymphovascular invasion (OR= 2.5). Importantly, the significance of certain variables weakened upon considering tumor location, underscoring its pivotal role as a covariate in evaluating FIT-IC-related factors. Although these findings provide important new insights, we could not extrapolate these observations to early stage CRC. To focus more on early stage detection of CRC, we evaluated DNA methylation patterns in colorectal

normal, precursor and carcinoma tissue. With this investigation into methylation patterns in colorectal tissues, we aimed to fill a critical knowledge gap. On one hand, an epigenome-wide study for these three tissue types was never performed simultaneous. On the other hand, it is clinically very relevant to differentiate between adenomas and carcinomas, since treatment and follow-up varies. Starting from a comprehensive dataset sourced from both 450K and EPIC publicly available datasets, we "double-evidenced" differentially methylated sites. Using previously developed computational methods, we continued the analysis to discover potential biomarkers. We identified 13 double-evidenced target sites, which we further validated using our in-house clinical carcinoma and adenoma samples. Our final model yielded a sensitivity of 96% at 95% specificity for distinguishing between adenomas and carcinomas. This study highlighted the use of methylation as a source for cancer biomarkers and provided us with potential early detection biomarkers for CRC.

In the third part of this thesis, we further explored the epigenomics field. Important in cancer diagnostics, is the observation that DNA methylation biomarkers hold promise for accurate detection across various tumor types. To harness this potential for multi-cancer detection, we developed a multiplex droplet digital PCR (ddPCR) assay. Drawing upon data from The Cancer Genome Atlas (TCGA), we selected differentially methylated targets for eight prevalent tumor types and validated them in tumor and normal adjacent samples. The resulting assay demonstrated high accuracy (AUC=0.948), albeit with varying sensitivities (53.8% - 100%) and specificities (80% -100%) across cancer types. Although we obtained promising results with ddPCR, we encountered the limitation that only a few targets can be simultaneously analyzed using this technology. Building upon the observation from our previous study, namely, 13 targets collectively capable of distinguishing adenomas from carcinomas, and considering the notion that multiple targets combined can enhance sensitivity, we have developed a novel technique called IMPRESS (Improved Methylation Profiling using Restriction Enzymes and smMIP Sequencing). By ingeniously combining methylation-sensitive restriction enzymes with single-molecule Molecular Inversion Probes, IMPRESS facilitates multiplex analysis without bisulfite conversion. Leveraging this technique, we successfully developed a multi-cancer detection assay for eight lethal cancer types, which we tested in 111 tumor and 114 normal fresh frozen samples. With our final classifier model, we achieved high sensitivity (95%) and specificity (91%) in distinguishing tumor from normal tissue. Lastly, our first preliminary results suggest that IMPRESS is very likely to perform well in cfDNA from liquid biopsies.

In the final part of this thesis, we aimed to create screening assays for CRC and breast cancer (BRCA) using IMPRESS. Using a similar strategy as in part two, we leveraged publicly available 450K array datasets to find differentially methylated sites for BRCA. We compared whole blood, normal, in situ and invasive carcinoma tissue to identify potential methylation biomarkers. For CRC, we partially re-used the already performed EPIC array analysis to select sites that were compatible with IMPRESS and re-performed some analysis with whole blood. The latter was used in view of applicability in liquid biopsies. A total of 152 smMIPs were used to analyze

9 BRCA, 9 in situ, 9 normal breast samples and 180 smMIPs were used for evaluation of 25 CRC, 38 adenoma and 30 normal colon samples.

For both CRC and BRCA, we created a two-step assay, where the first step assay aims to discriminate cancerous samples against normal samples (cancer detection panel), and the second step assay is used to distinguish precancerous lesions from invasive tumors (invasiveness detection panel). For CRC, cancer detection panel reached 100% sensitivity and 100% specificity. The invasiveness detection panel achieved 80% sensitivity at 92% specificity, with an accuracy of 0.88. For BRCA, the cancer detection panel had a sensitivity of 94.4% at 100% specificity. The accuracy of the invasiveness detection panel was 0.77, with a sensitivity of 66.7% at 88.9% specificity. Our preliminary results indicate high sensitivity and specificity for discriminating cancerous samples against normal samples and good accuracy for distinction between precancerous and invasive tumors. Further research will be needed to validate these panels. Lastly, provide a solid basis for examination of these panels in plasma-derived liquid biopsies.

The integration of innovative technologies, such as multiplex ddPCR assays and IMPRESS, into cancer diagnostics underscores the significance of epigenetic research in cancer biomarker discovery and clinical implementation. These advancements offer promising avenues for enhancing the accuracy and efficiency of tumor detection across various cancer types. As we continue unraveling the nuances of DNA methylation patterns, the results of this thesis pave the way for improved clinical applications in cancer screening and detection.

SAMENVATTING

Kanker blijft een belangrijk probleem in de wereldwijde gezondheidszorg, ondanks sterke vooruitgang in biomedisch onderzoek. Waar de meeste kankers gemakkelijk te behandelen zijn wanneer ze vroeg worden ontdekt, kan het vinden van de juiste therapieën uitdagend worden in latere stadia. Tijdige screening speelt dus een cruciale rol bij het verbeteren van de prognoses voor kankerpatiënten. Traditionele diagnostische methoden zoals lichamelijk onderzoek, beeldvorming en weefsel biopsieën zijn echter vaak invasief, duur en niet altijd nauwkeurig. Hoewel er verschillende innovatieve diagnostische technieken zijn ontwikkeld, blijft de zoektocht naar zowel betrouwbare biomerkers als gevoelige technologieën doorgaan. Binnen het biomerker onderzoek is DNA methylatie naar voren gekomen als een veelbelovende epigenetische factor met een significant impact in het kanker onderzoek. Daarom is onderzoek naar DNA methylatie als een potentieel hulpmiddel voor vroege detectie van kanker cruciaal.

Als respons op de groeiende belangstelling, is er aanzienlijke vooruitgang geboekt in het detecteren van DNA methylatie, met name in de context van vloeibare biopsieën en celvrij DNA (cfDNA). Traditionele methoden richtten zich op genomisch DNA, maar recente innovaties hebben zich verdiept in enzymatische methoden, vooral voor cfDNA-analyse. Dit heeft geleid tot een heel arsenaal aan zowel klassieke als nieuwe methoden voor de detectie van DNA methylatie, inclusief enkele klinische toepassingen gecertificeerd door de Europese *in vitro* diagnostiek wetgeving (CE-IVD). Opkomende trends houden mogelijke overgang naar enzymatische conversies in als de nieuwe standaard, alsook de integratie van directe sequentiemethoden zoals ONT-sequencing in epigenetisch onderzoek. Ten laatste is er de integratie van multi-omics technologieën, die beloftevol zijn voor verbeterde klinische toepassingen door diverse biomerkers te combineren voor verbeterde diagnostische nauwkeurigheid.

In het tweede deel van dit proefschrift hebben we ons gericht op de ontdekking van nieuwe biomerkers, met name bij colorectale kanker (CRC). In een eerste studie bestudeerden we het CRC-screeningsprogramma en de problemen ervan. In het bijzonder keken we naar intervalkankers (IC's), omdat ze een grote uitdaging vormen in CRC-screeningsinitiatieven. Eerdere onderzoeken door onze samenwerkingspartners wierpen al licht op IC's gedetecteerd door de fecale immunochemische test (FIT), waarbij hun prevalentie in specifieke demografische en tumorkenmerkende subsets werd benadrukt. Voortbouwend op deze inzichten, heeft onze studie zich gericht op stadium IV CRC-patiënten om factoren geassocieerd met FIT-IC te verhelderen. Door analyse van klinisch-pathologische en moleculaire kenmerken van 500 CRC's, hebben we associaties geïdentificeerd tussen de klinisch-pathologische en moleculaire kenmerken en FIT-IC stadium IV CRC. Specifiek vertoonden FIT-IC CRC's significant hogere kansen om neuro-endocriene tumoren (NET's) (OR= 5) te zijn en lymfovasculaire invasie te vertonen (OR= 2,5). Belangrijk is dat de significantie van bepaalde variabelen afnam bij het in rekening brengen

van de tumorlocatie, waarbij de essentiële rol ervan als covariabele bij de evaluatie van FIT-IC-gerelateerde factoren werd benadrukt. Hoewel deze bevindingen belangrijke nieuwe inzichten bieden, konden we deze observaties niet extrapoleren naar CRC in een vroeg stadium. Om ons meer te richten op de vroege detectie van CRC, hebben we DNA methylatiepatronen geëvalueerd in colorectaal normaal, precursor en carcinoomweefsel. Met dit onderzoek naar methylatiepatronen in colorectaal weefsel wilden we een belangrijke tekortkoming in het veld bekijken. Enerzijds werd een epigenoom-wijd onderzoek voor deze drie weefseltypes nog nooit simultaan uitgevoerd. Anderzijds is het klinisch zeer relevant om onderscheid te maken tussen adenomen en carcinomen, aangezien de behandeling en opvolging verschillen. Vertrekkend vanuit een uitgebreide dataset afkomstig van zowel 450K als EPIC openbaar beschikbare datasets, hebben we "dubbel-bewijsmateriaal" voor verschillend gemethyleerde locaties geïdentificeerd. Met behulp van eerder ontwikkelde computationele methoden hebben we de analyse voortgezet om potentiële biomerkers te ontdekken. We identificeerden 13 biomerkers, die we verder valideerden met onze eigen klinische carcinoom- en adenoom stalen. Ons uiteindelijke model leverde een sensitiviteit van 96% bij 95% specificiteit voor het onderscheid te makken tussen adenomen en carcinomen. Deze studie benadrukte het gebruik van methylatie als een bron voor kanker biomerkers en leverde ons potentiële vroege detectie biomerkers op voor CRC.

In het derde deel van dit proefschrift hebben we het epigenomisch veld verder verkend. Belangrijk in kanker diagnostiek, is de observatie dat DNA methylatie biomerkers beloftevol zijn voor nauwkeurige detectie over verschillende tumortypes. Om dit potentieel voor multi-kankerdetectie te benutten, hebben we een multiplex droplet-digital PCR (ddPCR) test ontwikkeld. Uit de gegevens van The Cancer Genome Atlas (TCGA), hebben we verschillend gemethyleerde doelwitten geselecteerd voor acht veelvoorkomende tumortypes en deze gevalideerd in tumor- en normale aangrenzende stalen. De resulterende test toonde een hoge nauwkeurigheid (AUC=0,948), zij het met variërende sensitiviteit (53,8% - 100%) en specificiteit (80% -100%) over tumortypes. Hoewel we veelbelovende resultaten behaalden met ddPCR, stootten we op de beperking dat slechts enkele doelwitten tegelijk kunnen worden geanalyseerd met deze technologie. Voortbouwend op de observatie uit ons vorige onderzoek, namelijk 13 CpGs gezamenlijk in staat tot het onderscheiden van adenomen van carcinomen, en rekening houdend met de hypothese dat meerdere biomerkers gecombineerd de gevoeligheid kunnen verbeteren, hebben we een nieuwe techniek ontwikkeld genaamd IMPRESS (Improved Methylation Profiling using Restriction Enzymes and smMIP Sequencing). Door methylatie-gevoelige restrictie-enzymen te combineren met single-molecule Molecular Inversion Probes, vergemakkelijkt IMPRESS multiplex analyse zonder bisulfiet conversie. Door deze techniek te benutten, hebben we succesvol een multi-kankerdetectie-test ontwikkeld voor acht dodelijke kankertypes, die we hebben getest in 111 tumor- en 114 normale vers ingevroren stalen. Met ons uiteindelijke classificatiemodel behaalden we een hoge sensitiviteit (95%) en specificiteit (91%) in het onderscheiden van tumorweefsel van normaal weefsel. Ten slotte suggereren onze

eerste voorlopige resultaten dat IMPRESS zeer waarschijnlijk goed zal presteren in cfDNA van vloeibare biopsieën.

In het laatste deel van dit proefschrift richtten we ons op het creëren van screeningsassays voor CRC en borst kanker (BRCA) met behulp van IMPRESS. Met een vergelijkbare strategie als in deel twee, hebben we gebruik gemaakt van openbaar beschikbare 450K array datasets om verschillend gemethyleerde locaties voor BRCA te vinden. We vergeleken volbloed, normaal, in situ en invasief carcinoomweefsel om potentiële methyleringsbiomerkers te identificeren. Voor CRC hebben we gedeeltelijk de reeds uitgevoerde EPIC array-analyse hergebruikt om locaties te selecteren die compatibel waren met IMPRESS, en hebben we enkele analyses opnieuw uitgevoerd met volbloed. Het laatste werd gebruikt met het oog op toepasbaarheid in vloeibare biopsieën. In totaal zijn 152 smMIPs gebruikt om 9 BRCA, 9 in situ en 9 normale borststalen te analyseren, en 180 smMIPs om 25 CRC, 38 adenoom en 30 normale colonstalen te evalueren.

Voor zowel CRC als BRCA hebben we een twee-staps process gebruikt, waarbij de eerste stap erop gericht is om tumoren te onderscheiden van normale stalen (kanker detectie panel), en de tweede stap assay wordt gebruikt om precancereuze laesies van invasieve tumoren te onderscheiden (invasiviteit detectie panel). Voor CRC bereikte het kanker detectie panel een sensitiviteit van 100% en een specificiteit van 100%. Het invasiviteit detectie panel had een sensitiviteit van 80% bij 92% specificiteit, met een AUC van 0,88. Voor BRCA had het kanker detectie panel een sensitiviteit van 94.4% bij een specificiteit van 100%. De AUC van het invasiviteit detectie panel was 0,77, met een sensitiviteit van 66,7% en een specificiteit van 88,9%. Onze voorlopige resultaten geven aan dat er een hoge sensitiviteit en specificiteit is voor het onderscheiden van tumoren en normale stalen en een goede nauwkeurigheid voor het onderscheid tussen precancereuze en invasieve tumoren. Verder onderzoek zal nodig zijn om deze panels te valideren. Tot slot bieden we een mooie basis voor het onderzoeken van deze panels in plasma-afgeleide vloeistofbiopten

De integratie van innovatieve technologieën, zoals multiplex ddPCR-testen en IMPRESS, in kanker diagnostiek benadrukt het belang van epigenetisch onderzoek in de ontdekking en klinische toepassing van kankerbiomerkers. Deze vooruitgang biedt veelbelovende mogelijkheden om de nauwkeurigheid en efficiëntie van tumordetectie over verschillende kankertypes te verbeteren. Terwijl we doorgaan met het ontrafelen van de nuances van DNA methylatie patronen, leggen de resultaten van dit proefschrift de weg vrij voor verbeterde klinische toepassingen in kankerscreening en - detectie.

PART I | Introduction

CHAPTER 1

From Classic to Recent DNA Methylation Detection Methods in Cancer: The Current Technology Landscape

(Manuscript submitted)

Isabelle Neefs[1,2], Joe Ibrahim[1,2], Marc Peeters[2], Guy Van Camp[1,2] and Ken Op de Beeck[1,2]

1 Centre of Medical Genetics, University of Antwerp and Antwerp University Hospital, Prins Boudewijnlaan 43, 2650 Edegem, Belgium

2 Centre for Oncological Research Antwerp (CORE), University of Antwerp and Antwerp University Hospital, Universiteitsplein 1, 2610 Wilrijk, Belgium

ABSTRACT

Background: DNA methylation is a well-studied epigenetic factor and has become a powerful player in the cancer biomarker research field. Together with the rising interest in methylation biomarkers, the technological advances for the detection of DNA methylation have been immense. This has led to a plethora of different methods. The first methods were established for DNA methylation detection in genomic DNA, while new methods have focused more on compatibility with the emerging interest of cell free DNA (cfDNA) from liquid biopsies. As DNA methylation detection in cfDNA brings its own challenges, a shift from the gold standard bisulfite conversion towards enzymatic conversion methods can be observed in recent years.

Main body: In this review, we aim to summarize the classic and more recent DNA methylation detection methods. Importantly, the few existing European Certified in vitro diagnostics (CE-IVD) clinical applications for liquid biopsies are also described, underlining the potential of DNA methylation as a detection biomarker in cfDNA. Furthermore, we provide some insights into how the field might evolve in the future, where we believe enzymatic conversions might become a new gold standard and direct sequencing methods, such as ONT-sequencing, will get an important place in the epigenetic research field. Lastly, we believe that multi-omics technologies, which can combine distinct types of biomarkers, will most likely become more important in future clinical applications.

Conclusions: The few emerging CE-IVD assays indicate that methylation could become an important part of clinical diagnostics. Moreover, novel recent technologies are being developed and show promising clinical applications. Taken together, methylation biomarkers are becoming more important for clinical implementation.

INTRODUCTION

One of the most studied epigenetic alterations is DNA methylation. It is the addition of a methyl group (-CH_3) to the fifth position of a cytosine in a CpG dinucleotide context. CpG dinucleotides only represent about 1% of the mammalian genome but tend to cluster in so-called CpG islands (CGI). CGIs occur in the majority of human protein-coding genes near the transcription start site but can also occur in the gene body [1,2].

Methylation is crucial for normal cell functioning and development, so abnormalities in the process can lead to various diseases. Not only does its absence or presence mediate its biological function, but also its location in the genome [3]. Methylation changes occur early in cancer development and typically, promotor regions of genes (e.g. tumor suppressor genes) are hypermethylated, while the gene body and repetitive elements are hypomethylated. This leads to genomic instability and dysregulation of normal cell function [2,4].

Initially, methylation studies focused on the promotor and transcription start site (TSS) hypermethylation of protein-coding genes. However, more recent genome-wide analyses revealed the role of methylation at CpG sites within introns, intergenic sequences and exons, which led to the identification of CpG shores (2kb from CGI), CG shelves (2-4kb from CGI) and open sea regions (rest of the genome) [1]. This shows the enormous potential of the epigenome as a source of methylation biomarkers which are already used for several diseases, including neurodegenerative disorders, cardiovascular diseases, and cancer [3]. The use of methylation markers for disease detection has several advantages over mutation-based detection. Methylation is an early event in disease development, so it is very interesting for early detection. Also, methylation signatures are more universal than mutation markers, which typically vary at a wide range of sites. Given that no prior knowledge is needed on the tumor molecular profile, methylation-based tests can be used off-the-shelf, making them much faster and cheaper to use [5,6]. Lastly, methylation can be used to study the tissue of origin, which is of great importance when biomarkers are detected in tissue or fluids taken from another origin [7]. Methylation could potentially be used to monitor many common diseases with a simple cfDNA blood test by taking advantage of these tissue-specific differences [8]. With the increased interest in detecting methylation, technological advances have been enormous. Multiple methods are now available for methylation studies. However, most of the technologies have important issues such as high cost, inapplicability for some sample types, high DNA input requirements, complex (bioinformatic) analyses and low sensitivity and/or specificity. These issues are even more pronounced in a liquid biopsy context, where (cf)DNA is often already fragmented and available in limited amounts [9]. Nevertheless, several studies on cfDNA methylation detection have been performed.

Over the years, many methylation detection techniques have been developed. This manuscript provides an overview of both classic and more recent techniques along with their (dis)advantages (see Table 1). We restrict

this review to the most used technologies (see Figure 1). Furthermore, the current use of methylation biomarkers in liquid biopsies in the clinic is described.

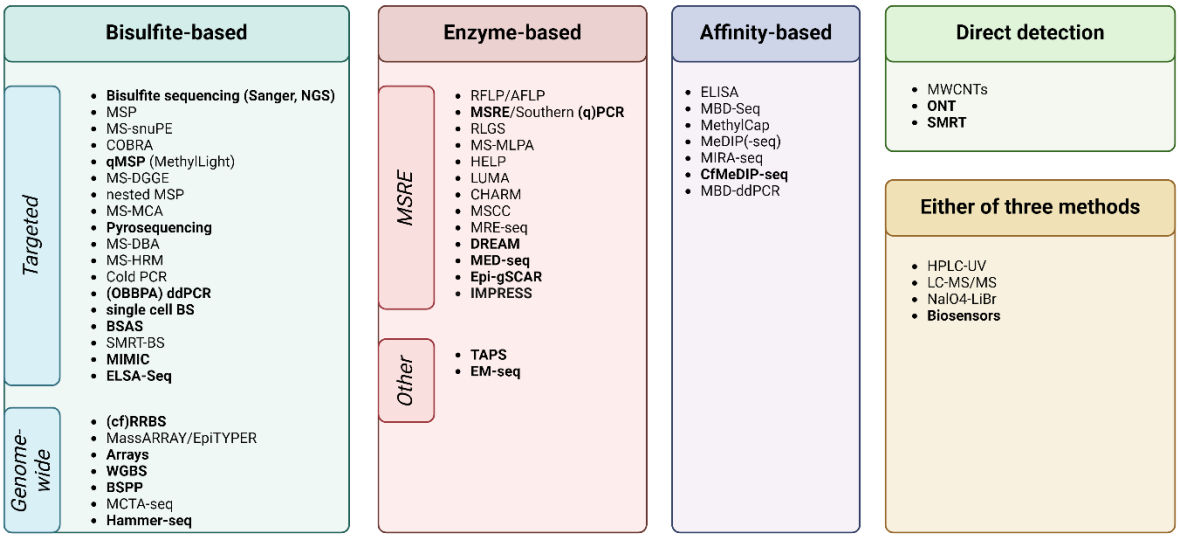


Figure 1 | Overview of DNA methylation detection technologies. Methods in bold are the ones described in detail in this article. More information can also be found in Table 1.

CLASSIC METHYLATION METHODS

DNA methylation plays a crucial role in several biological processes. Methylated (5mC) and non-methylated (C) cytosine residues can be discriminated using various techniques. These can be grouped per methylation detection strategy: 1) bisulfite conversion-based methods, 2) restriction enzyme-based assays and 3) affinity enrichment-based approaches. In recent years, direct detection of methylation has gained a lot of interest (Fig. 1). This review gives an overview of classic and novel methylation detection technologies and their implementation in the clinic.

Bisulfite-based methods

Bisulfite conversion can be used to determine the methylation status of a single, specific locus or a more general profile in a wider region. Therefore, techniques are divided into two categories: target or locus-specific methods and genome-wide approaches [10–20]. All bisulfite-based methods share limitations inherent to bisulfite treatment. Bisulfite conversion is a harsh chemical method (high pH and temperature) that degrades and fragments DNA, resulting in a poor-quality product and loss of input DNA. This is mostly a disadvantage in samples where a limited amount of DNA is available [12,13,16–18,20–22]. Related thereto, significant degradation of the bisulfite-treated DNA can occur upon storage, as the single-stranded DNA is unstable after conversion. Therefore, converted DNA must be analyzed shortly after conversion to not impair the sensitivity

of downstream applications [18,23]. Furthermore, incomplete conversion of cytosine to uracil, for example by re-annealing of ssDNA or by the presence of residual RNA or proteins, can introduce artifacts and lead to inconsistencies in results [11,12,17,18,21,23]. Although commercially available kits are highly efficient (>97% to 99%), the conversion rate can vary slightly between one another [17,19,20]. Moreover, bisulfite treatment cannot discriminate between 5mC and 5hmC, which could affect downstream analyses [12,17,19,23].

Target-specific bisulfite-based methods

(Targeted) bisulfite sequencing

Since the development of bisulfite sequencing in 1992, many efforts have been invested to make the use of bisulfite more time and cost-effective and to lower the DNA input, which was a major problem in the original protocol [24]. For example, Masser *et al.* developed the Bisulfite Amplicon Sequencing (BSAS) method. It combines elements such as the bisulfite treatment and region-specific PCR amplification with transposome-mediated NGS library construction and benchtop NGS. However, it is still bisulfite-based and expensive for multiplexing more than 20 genomic targets [22,25]. In other targeted bisulfite sequencing methods, enrichment of CpG-rich regions or specific regions of interest (ROIs) is performed based on hybridization probes, before or after bisulfite treatment and followed by NGS [23,26]. Capture of the DNA before bisulfite conversion ensures better enrichment since the genome is not yet converted but requires more DNA. Hybridization of the probes after bisulfite conversion produces more off-target reads but results in higher complexity of the sequencing library [23]. Several kits for specific targets (disease panels) or regulatory regions are commercially available. Examples include the Agilent Sure-Select Methyl-Seq and TruSeq Methyl Capture panels (capture before bisulfite conversion) and the Roche SeqCap Epi (hybridization after bisulfite treatment) [22,23,26].

With the rise of interest in single cell analysis around 2010, the development of methylation-specific single-cell analysis methods gained interest. Degradation due to bisulfite treatment and limited DNA input capacity is circumvented in single-cell BS with post bisulfite adaptor tagging (PBAT). This further allows PCR amplification and deep sequencing at the single-cell level, although PBAT comes with biases and the formation of chimeric reads [15,22]. Because of random fragmentation and random priming, a single-cell library covers only around 10-20% of the genome.

An application of bisulfite (NGS) sequencing is the IvyGene Cancer Blood test (Laboratory for Advanced Medicine Inc., Irvine, CA, USA). This test can detect multiple cancer types, including breast, colon, lung and liver cancer, simultaneously. It is currently only available as laboratory developed test (LDT) and targets the methylation status of the *MYO1G* and *TNFAIP8L2* genes in cfDNA. As preliminary results show that a positive result can still indicate another cancer type, the current IvyGene test must be used in combination with other diagnostic tests [27].

Methylation-specific (q)PCR and related technologies

Methylation-Specific PCR (MSP) was described in 1999 but became somewhat obsolete with the introduction of qPCR. MSP and related technologies are described in supplementary table 1. To obtain quantitative information, real time MSP (qMSP) protocols were developed [13]. An intercalating dye such as SYBR green (MethylQuant), or methylation specific TaqMan probes can be used in the so-called MethyLight assay [11,18,20]. Furthermore, qMSP eliminated the use of gels, leading to its implementation in the clinic [17–19]. qMSP is ten times more sensitive than MSP and can detect low frequencies of hypermethylated CpGs (0.01%) [12,17,28]. In both MSP and qMSP (related) methods, two major disadvantages include the lack of multiplexing possibilities, reducing sensitivity in fragmented samples, and the lack of standardization amongst different labs [11,26]. Only the multiplex MethyLight assay can theoretically analyze multiple genes simultaneously [12], but in practice, no more than three targets are evaluated simultaneously [29–31]. More related technologies are described in Table 1 and Supplementary Table 1. Quite a few applications have been FDA approved, amongst them the Epi proColon 2.0 CE (Epigenomics AG, Berlin Germany). This blood-based test uses a MethyLight assay to detect methylated *SEPT9* aiming to improve early diagnosis of colorectal cancer. Across several studies, this test has been shown to discriminate healthy controls and CRC patients with a sensitivity of 75-81% and a specificity of 96-99% [28]. Although it was approved in 2016, European guidelines still do not recommend this test to be used as a first choice in screening programs [27,32]. Cost-effectiveness analyses, further validation and regulatory approval will be needed to implement the EpiProcolon in programs across Europe.

Pyrosequencing

Despite being an older method, pyrosequencing remains an important commercial technology, for which several kits, platforms and software from different companies are available today [19,26,28]. The technique itself is described in detail elsewhere [11,12,17–20,28,33,34] and its (dis)advantages are listed in Table 1.

One of its main commercial applications are the different kits for determining the methylation status of the *MGMT* gene, such as the Therascreen MGMT Pyro kit (Qiagen, Hilden, Germany). *MGMT* is a known biomarker for glioblastoma classification and treatment decision-making. Pyrosequencing was found to lead to the most reproducible results, hence the choice for this technology in the commercially available kits. All kits are commercialized by Qiagen, but different systems (e.g. the PyroMark Q24 for the Therascreen kit) are used with different kits, and the Therascreen is the only kit that obtained the CE-IVD mark. Using this kit, 4 CpG sites in the *MGMT* gene are analyzed in DNA obtained from either blood or FFPE samples [27,32].

Droplet Digital PCR

Droplet digital PCR (ddPCR) was developed to increase the analytical sensitivity of PCR for detecting rare events, for example methylation of a specific allele. Bisulfite converted DNA is fragmented into thousands of droplets, allowing as many simultaneous PCR reactions. In theory, each droplet contains only one DNA template, which is negative or positive for the rare event. The fraction of positive droplets represents the rare event. The enormous sample partitioning allows quick amplification, and screening of separate DNA molecules with a limit of detection up to 0.001% and absolute quantification. However, primer design is laborious and difficult [11,12,17]. Furthermore, the multiplex capacities for ddPCR are very limited, with at most five targets reported in one assay [35]. To further increase the sensitivity, Menschikowski *et al*. developed optimized bias-based pre-amplification ddPCR (OBBPA ddPCR) where they reached analytical sensitivities of 0.0007%. However, the procedure cannot be performed in a close-tube system, increasing sample contamination risks. The technology is claimed to be especially suitable in liquid biopsies [36,37]. Although ddPCR is a very promising method, there are currently no ddPCR methylation-based applications FDA-approved for clinical use.

2.1.1.5 MIMIC

Schwalbe *et al*. developed the minimal methylation classifier (MIMIC) for the assessment of 17 CpG sites simultaneously. Bisulfite converted DNA is subjected to a single base extension of probe oligonucleotides, followed by MALDI-TOF MS. Although low input DNA (< 2ng) can be used and an accurate assessment of methylation level is made, this technology is expensive [38].

Genome-wide bisulfite-based methods

WGBS

Since the emergence of next-generation sequencing (NGS) technologies, combining bisulfite conversion with NGS has become the new gold standard for global methylation analysis. Using Whole Genome Bisulfite Sequencing (WGBS) analysis, information about the methylation status of every cytosine in the methylome can be obtained, including low-density regions. Relatively low coverage (5-10x) is often used for highly reproducible and accurate determination of methylomes. Although it is the most comprehensive methylation profiling technology, there are some disadvantages (see also Table 1). WGBS is relatively expensive and as most of the human methylome is not methylated, a lot of sequencing capacity and money is wasted [10,13,15,19,23,24,39]. However, sequencing costs have been decreasing over the years, potentially making WGBS economically more feasible in the future [13]. Furthermore, due to the reduction of genome complexity and the loss of sequence diversity after bisulfite conversion, bioinformatic analysis of the WGBS data is difficult [16,22,26,39]. Specific pipelines, such as Bismark, gemBS and Methylpy, have been designed to streamline the

pre-processing of this data and to facilitate the highly complex WGBS alignment. Novel ones, such as Methyldackel and MethylStar, are still emerging [40]. However, bioinformatic expertise and computational resources remain necessary [39].

RRBS

Before the introduction of WGBS, Reduced Representation Bisulfite Sequencing (RRBS) was already used in epigenetic DNA analysis. For this technology, methylated regions are enriched using *MspI* digestion (at CCGG sites) and bisulfite conversion before NGS. Digestion and size selection (40-220 bp) allow for the assessment of a smaller fraction of the genome while still allowing representative results [13,22,24]. Around 85% of CpG islands are evaluated with this technology, which only represents <3% of the genome [13,15,22,26]. As such, the cost is drastically decreased compared to WGBS, while the sample throughput is increased [13,18,24]. Despite RRBS being more reproducible than affinity-based methods for example, it still is less reproducible than WGBS and arrays due to its use of enzymes, targeted nature and lower sequencing depth [18,19,41]. Moreover, CpG-rich regions are enriched, while distal regulatory elements and intergenic regions have relatively low coverage in RRBS as the CpG-containing recognition sites for *MspI* are limited [13,18,23]. To sequence CpG island shores, longer restriction fragments can be sequenced in a technique known as enhanced RRBS [22]. In recent years, single-cell RRBS has been developed to enable the use of this technology in cfDNA. DNA loss is avoided by integrating all key steps in a single-tube reaction [13,42]. However, since adaptor hybridization takes place before bisulfite conversion, a fraction of well-ligated fragments is destroyed by the bisulfite and is therefore unavailable for amplification in library preparation [43]. Both RRBS and WGBS protocols have been adapted for use in single-cell BS [43].

cfRRBS

Recently, RRBS was adapted for use in cfDNA by De Koker *et al.* [44]. They found an ingenious way to circumvent the problem cfDNA fragmentation, which hampers RRBS applicability. With their cfRRBS protocol, only MspI generated fragments are amplified, while “off-target” cfDNA fragments are degraded. This is possible due to the use of hairpin adaptors that specifically bind to the phosphorylated 5’-ends that are created at MspI cut sites. These circular fragments withstand the exonuclease treatment. Thereafter, classical RRBS libraries are generated using bisulfite conversion and PCR amplification [44]. Some interesting applications have already been published, where cfRRBS is used to develop a diagnostic test for Cancers of Unknown Primary (CUP)[45] and pediatric cancers [46–48]. It is a cost-effective, scalable method and thanks to the single-tube protocol, it allows the use of as little as 0.4 - 10 ng DNA input [44,45]. However, despite covering 3 million CpG sites, targets remain limited to MspI restriction sites, which might exclude interesting candidate-biomarkers.

Arrays

Throughout the last decades, Illumina has brought several array types to the market. The first-generation assays were replaced by the HumanMethylation450 BeadChip in 2011 [49]. It was used for large projects such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC). In 2016, the novel methylation EPIC array was released [50]. This assay interrogates over 850,000 preselected CpG sites, covering 90% of 450K sites and more sites within enhancer regions [13,19]. As of 2023, the Infinium Methylation EPIC v2.0 has replaced the former version. Compared to the v1, poor-performing probes have been removed and an additional 186,000 CpG sites have been added. They target enhancers, CNV detection regions and additional CTCF-binding sites. Moreover, >450 cancer driver mutations were added. Importantly the v2.0 has been validated for use in FFPE material [51,52]. All arrays work with similar technology, described elsewhere [19,26]. Methylation arrays are relatively low-cost techniques, give accurate measurements and are suitable for large sample numbers [39,50]. However, cross-hybridization of probes, erroneous signals from SNPs and probe-specific dye biases lead to some bioinformatic challenges. Moreover, batch effects can occur if not handled properly. Also, a large input amount is still required (500-1µg) [17,23,39].

Based on the HumanMethylation450 BeadChip, the EPICUP™ test (Ferrer, Spain) was developed for the biological defining of tissue of origin in Cancer of Unknown Primary (CUP) in FFPE or fresh frozen samples. By comparing methylation patterns of known primary cancers with CUP, the origin can be predicted with high sensitivity and specificity. The 450K chips are no longer manufactured, but recent studies show identical results with the Infinium EPIC arrays. Despite the importance of a CUP test, the EPICUP™ test cannot be executed in all laboratories due to the technology behind it, which leads to rather long TATs of 2 weeks [27,32,53].

ELSA-seq

Liang *et al*. recently described the 'Enhanced Linear-Splinter Amplification Sequencing (ELSA-seq)', an improved bisulfite-based method for cfDNA. With ELSA-seq, detection power is on the one hand improved by increasing the DNA template that can be effectively used. WGBS libraries are generated and a dual-index system is used for multiplexing samples. On the other hand, detection power is increased by the use of machine learning. This high-resolution technology allows for very low input DNA (500 pg) and has very effective noise suppression, but the adaptor tagging step may lead to incomplete duplicate removal. This method has only been validated in lung cancer up until now, but it is already a promising technology for cfDNA applications [54]. The OverC™ test for multi-cancer detection was developed based on this technology. It received the FDA breakthrough device designation, but it is not described as an LDT nor is there FDA approval or CE-IVD for this test [53].

Hammer-Seq

Among the most recent bisulfite-based methods, Ming *et al.* introduced "Hairpin-Assisted Mapping of Methylation of Replicated DNA Sequencing", hammer-seq in short in 2021. This technology combines EdU labeling of replicated DNA, biotin-streptavidin-based purification and whole-genome hairpin bisulfite sequencing technologies. Their method's most unique feature is the simultaneous measurement of methylation on both parent and daughter strands. The technique is ideal for determining maintenance kinetics and *de novo* methylation events occurring during methylation maintenance. However, there is no way to determine which strand is the parent strand and which one is the daughter strand. Moreover, it requires around 100 µg genomic DNA, which renders it unfeasible for applications with lower input DNA for the moment [55].

Enzyme-based methods

Restriction enzymes

The use of restriction enzymes was the first approach to assess locus-specific methylation [3,10,26]. Two types of enzymes are mostly used. Methylation-sensitive restriction enzymes (MSREs, e.g. HpaII) are inhibited by the presence of methylation at their recognition site, reflecting the distinct methylation status at a specific locus [10,13,24]. Methylation-insensitive restriction enzymes (e.g. MspI) cleave the DNA regardless of the methylation status at the recognition sites [13,42]. To distinguish methylated and unmethylated CpGs, pairs of both enzymes with the same restriction site but different sensitivity to methylation status (isoschizomers) can be used [12,14]. Depending on the application and technology, MSREs with recognition sites between 4 to 8bp have been used [56]. A great advantage is that the need for bisulfite conversion is eliminated using these enzymes, which in turn leads to lower DNA input and easier primer design [18,20]. It has also been described that restriction-based methods are superior to other enrichment technologies as a conversion method prior to NGS [26]. They are fast, specific and easy to use [20]. However, only loci with the restriction site(s) of the enzyme(s) can be investigated, and upon incomplete digestion, false-positive results are a possibility [11,18] (see also Table 1).

MSRE-(q)PCR

MSRE-based methylation assays were traditionally used in combination with Southern blotting, but then switched to PCR and later qPCR. With this, DNA input drastically decreased, as well as the cost and simplicity.

Combining MSREs with qPCR was first done in 2005. Primers flanking the region of interest are used to analyze methylation. Methylation percentages are counted from the Ct values that are measured for both digested and undigested control samples. Commercial kits have been developed to target multiple sites (e.g. OneStep qMethyl kit from Zymo Research). A difficulty with this approach is that at least two restriction sites are required to be inside the amplicon, so it is not possible to investigate one particular CpG site [11,18,20,26].

Based on MSRE-qPCR technology, the Bladder EpiCheck (Nucleix Ltd) is a CE-marked test available for the detection of non-muscle invasive bladder cancer (NMBIC) recurrence. 15 methylation biomarkers are investigated in DNA extracted from urine. Studies have shown that the Bladder EpiCheck can be combined with cytology to reduce the invasiveness of NMBIC follow-up [27,53].

DREAM

Digital Restriction Enzyme Analysis of Methylation, DREAM in short makes use of SmaI and XmaI to create specific signatures at (un)methylated CpG sites. SmaI is used for digesting unmethylated CpG sites. XmaI has the same recognition site but cuts only at methylated sites. The two enzymes are used sequentially, so XmaI digests the remaining sites that have been protected from SmaI. Only fragments with these distinct signatures are sequenced and methylation levels are thus calculated. According to the authors, a total of 50,000 unique CpG sites are yielded with high coverage when sequencing 25 million reads per human DNA library. The background is reported to be less than 1%, making the technology suitable for low methylation level detection. However, 1µg of high-quality DNA is needed, hampering its applicability in liquid biopsy research [57,58].

MED-seq

Methylated DNA sequencing, MED-seq in short, was published by Boers *et al.* in 2018. In this technique, the methylation-dependent restriction enzyme LpnPI is used. LpnPI cuts 16 bp downstream from (hydroxy)methylated CpGs, leading to fragments of ⩾ 32 bp. This allows accurate identification of methylation. Although this method is relatively low-cost, simple and does not need much DNA input (10ng), only about 50% of methylated CpGs genome-wide can be detected. Unmethylated regions cannot be identified. Therefore, there is no possibility to completely quantify the methylation levels by a ratio or percentage, making this method only semi-quantitative [59]. Importantly, the technology has been successfully used in cfDNA samples, but only when 1) digestion with LpnPI was complete, and when 2) there was sufficient library-prepped DNA available. Starting concentrations were therefore 10ng. The MED-seq technology is described to be compatible with vacuum concentration, different blood collection tubes and cfDNA isolation methods [60].

EpiGScar

In 2021, EpiGScar was published by Niemöller *et al.* EpiGScar stands for Epigenomics and Genomics of Single cells analyzed by restriction and allows simultaneous analysis of methylation and genetic variants of the same cell at base pair resolution. Because of its single-tube workflow, contamination risks and more importantly DNA loss are reduced. The method is described at length in the paper of Niemöller *et al.* They use the commonly described MSRE HhaI in combination with other enzymes, resulting in intact (methylated) or scar-tagged HhaI

(unmethylated) sites. The resulting PCR amplified library is sequenced using NGS. Despite its good genome coverage (1.69 million RE sites), this technique is quite expensive and laborious [61].

Improved Methylation Profiling using Restriction Enzymes and SmMIP Sequencing (IMPRESS)

Very recently, a novel methylation detection method, called Improved Methylation Profiling using Restriction Enzymes and SmMIP Sequencing (IMPRESS) was developed [62]. For this technology, a specific combination of MSREs and single molecule Molecular Inversion Probes (smMIPs), followed by NGS, was derived. Four MSREs (HinP1I, AciI, HpaII and HpyCH4IV) are used to cover around 40% of the epigenome. The advantages of these MSREs include the reliability, simplicity and cost-effectiveness of the digestion reaction, while the smMIPs allow for highly accurate capturing of the DNA sequence of interest. However, IMPRESS cannot detect both unmethylated and methylated CpGs. Furthermore, it covers 40% of the (epi)genome, limiting the biomarkers that can be investigated using this technology. Due to its recent addition to the methylation detection field, its applicability in the clinic remains to be proven.

Other enzymes

TAPS

In the past decade, novel enzymes for methylation detection have gained attention. For example, Liu *et al*. described the novel TAPS, short for TET-Assisted Pyridine borane Sequencing in 2019 [63]. The Tet methylcytosine dioxygenase 1 (TET1) enzyme oxidizes (hydroxy)methylated cytosines (5(h)mC) to carboxyl-cytosines (5caC) or formyl-cytosines (5fC). The novelty of TAPS lies in the subsequent 5caC/5fC-to-T transition chemistry. Pyridine borane is used to reduce 5caC to dihydrouracil (DHU), which is in turn converted to thymine by PCR. As such, cytosine modifications can be detected. An advantage of the TAPS technology is the possibility to also detect hydroxy-methylated cytosines. For this, 5hmC is glucosylated to 5gmC before the TET1 oxidation [16,63,64]. Compared to whole genome bisulfite sequencing, TAPS has the advantage that the four-base genome is preserved, which allows efficient alignment and primer design. However, borane-mediated conversion requires long incubation times under acidic conditions, although it is less destructive than bisulfite deamination [64].

EM-seq

New England Biolabs commercialized Enzymatic Methyl sequencing (EM-seq) in 2020 [65]. The complete method is purely based on enzymatic conversions. In a first conversion step, two sets of enzymes are used. DNA is treated with TET2 and/or T4-BGT. T4-BGT protects 5hmC, TET2 protects 5mC. Subsequently, APOBEC3A is used for the deamination of cytosines, but not the protected ones. PCR amplification allows distinction

between unprotected C's, that are converted to T's, and protected C-derivates (read as C). EM-seq combines the oxidation and deamination reactions with NEBNext library preparation. It has been shown that inputs down to 10 ng can be effectively used. Furthermore, EM-seq does not cause DNA damage or DNA fragmentation. EM-seq can also be combined with long-read platforms such as Nanopore. However, EM-seq is still quite expensive [16,64–66]. It has also been demonstrated that the TET2-enzyme favors certain DNA motifs (e.g. A upstream of CG) due to its intrinsic sequence specificity. The CG flanking sequences affect the TET enzyme conformation, which influences the TET function [67,68]. This is an important issue that must be carefully evaluated, as it might hamper EM-seq's clinical applications.

Affinity-based methods

Affinity-based methods make use of immunoprecipitation with antibodies specific for 5mC or affinity purification with methyl-CpG binding domain (MBD) proteins to enrich methylated regions for further analysis [11,13,18] (Figure 1, Table 1). Previously, combinations with array hybridization were mostly used, but a shift towards NGS is now observed [10]. Compared to bisulfite-based methods, the biggest advantages are the avoidance of cytosine conversion and the possibility to discriminate 5-hydroxymethylation from 5mC. Affinity-based methods do not require high-quality DNA [26], but their accuracy suffers compared to RRBS and WGBS. Their resolution is only around 100-300 bp, which does not allow for single CpG site studies, and are biased towards hypermethylated regions [13,42]. Moreover, absolute quantification is not possible using an affinity-based method [39]. Furthermore, the standard protocols require a large amount of DNA input, so optimization for cfDNA is necessary [13,42] (Table 1).

(cf)MeDIP

Methylated DNA immunoprecipitation or MeDIP uses, antibodies against 5mC to isolate the DNA fragment containing this modification, independently of the surrounding DNA sequence [24,69]. After shearing and immunoprecipitation, the isolated regions are further investigated by qPCR, array or NGS [24,70]. MeDIP works on single-stranded DNA (ssDNA), allowing the profiling of hemi-methylated sites [39]. In contrast to bisulfite sequencing, it is biased towards low-density CpG sites [69]. One of the most important limitations of this technology was the high DNA amount required. Recently, the protocol has been adapted by Shen *et al*. [71] for use in cell-free DNA. This cfMeDIP-seq technology makes use of exogenous lambda DNA as filler DNA to increase the initial DNA input. As such, cfDNA amounts of as little as 1ng can be used. However, the technology has not yet been validated with a sufficient number of independent clinical samples.

DIRECT DETECTION METHODS

Around 2010, Third-Generation Sequencing (TGS) and Multi-Walled Carbon Nano Tubes (MWCNTs) entered the market. These methods allow for native DNA sequencing based on electrochemical signals. TGS focuses on current-based signals and kinetics [22], while MWCNTs detect the DNA nucleotides based on their unique oxidation signals [14,72]. These electrochemical technologies are base-specific and can be used for rapid detection of bases without using prior conversion steps [22,72]. The focus in recent research has mostly been on TGS, which will be discussed further below.

Third-generation/long-read sequencing

A few years after the introduction of TGS, also called long-read sequencing, applications for methylation detection with these platforms arose. Currently, two technologies are available: Oxford Nanopore Technologies (ONT) and Single Molecule Real-Time (SMRT) sequencing from Pacific Biosciences (PacBio). Although the first TGS analyses for methylation detection were performed after bisulfite conversion, the strength of these platforms lies in sequencing native DNA. Base modifications are derived based on raw signals that are generated. Despite the advantage of eliminating the conversion step, some limitations still need to be overcome before these technologies can be applied in the clinic. For example, DNA is not amplified and thus requires a large input amount (1µg for ONT, 5µg for PacBio). Such amounts of DNA are often not available namely in FFPE or cfDNA samples. Moreover, single-base pair calling is not very accurate in long-read sequencing. Also, long-read sequencing is even more expensive than NGS. Lastly, the cumbersome data output ('big data') of these technologies brings along bioinformatic challenges [22,70,73].

ONT sequencing

ONT-sequencing is based on measuring the ionic current changes of native single-stranded DNA that is passed through a nanopore. Every base and its modification gives a unique signal that is analyzed using a trained artificial neural network for base-calling, such as Guppy [27,39,73,74]. However, there is no standardized base calling pipeline and the error rate varies from 5%-20%, depending on sequencing context [18,39]. Nevertheless, software is quickly evolving and with time, the error rate will improve.

SMRT sequencing

Single molecule real-time or SMRT sequencing from PacBio detects methylation by monitoring the polymerase kinetics, at high (150-250x) coverage [75,76]. The kinetics of the polymerase are followed as it synthesizes circular DNA double strands using different fluorescently labeled nucleotides. As such, both the nucleotide

sequence and the major modifications such as 5(h)mC can be detected simultaneously [15,18,22,39]. Currently, basecalling is done in CCS [74]. However, 5mC signals are relatively subtle, leading to lower detection sensitivities compared to e.g. 6mA modifications of bacterial genomes. Therefore, SMRT has only been used extensively to study bacterial genomes and bacterial methylations [15,18,22,39].

FUTURE PERSPECTIVES

The methylation detection field is rapidly evolving. Besides new methods, specifically designed for methylation analysis, some new applications have recently gained interest in epigenetic research after decades of genetic research. Well-established technologies such as liquid chromatography and mass-spectrometry, for example, can be used for methylation detection, but still require DNA conversion with one of the above-described methods, depending on downstream applications and the nature of the available sample [26,77–80] (Figure 1 and Table 1). A newer example is the DNA biosensor, the surface of which contains specific complementary probes that capture DNA. Signal transducers are then used to convert the recognition event into either an optical or an electrochemical signal. For the detection of methylated DNA, a conversion method (discussed above) is used first. A biosensor system provides high sensitivity and specificity, and low cost [19]. Despite their advantages, biosensor application on clinical samples is currently lacking, making it difficult to evaluate their implementation prospects in the clinic [18].

Methylation detection strategies vary given the research question. For now, bisulfite treatment remains the most used conversion method in methylation detection, either in a targeted (such as ddPCR) or a whole-genome method (including WGBS and arrays). However, a shift towards enzymatic methods has been observed in recent years. For example, MED-seq, TAPS and EM-seq have been developed and the latter was even commercialized. We believe enzymatic conversion for methylation detection will become more important in the future, as it is especially useful for analyzing liquid biopsies that have been gaining more attention. Due to their relatively low cost and ease of implementation, these methods can become useful in clinical diagnostics. One example is the newly developed and validated IMPRESS assay which targets multiple sites and is an inexpensive technique when combining many samples, but is pending further validation before being implemented in the clinic. Furthermore, direct detection methods such as ONT and SMRT-sequencing are relatively underexplored methods that could gain more interest in the future. Despite their great advantage of sequencing native DNA, improving error rates [74] and potential for cfDNA sequencing [81], these technologies will remain challenging to implement in the clinic due to cost and logistics. Lastly, combining different ‘omics into one assay, for example genomics and epigenomics, has been gaining attention in recent years. These multi-omics strategies require their own methods and will become important for future applications.

Table 1 | Overview of most important and recently developed methylation detection technologies by conversion approach

Approach	Detection	Technology	Advantages	Disadvantages	FDA/CE-approved tests and LDTs	References
Bisulfite-based	*Targeted*	Bisulfite sequencing	- Gold standard - Quantitatively study locus-specific methylation	**Linked to bisulfite treatment:** - DNA degradation (also upon storage)→ loss of initial DNA input - Fragmentation - Loss of information - Incomplete conversion leading to false-positive results - No detection of hydroxymethylation **Linked to sequencing technology** - Biased PCR amplification - Very low detection limit (10-20%)	- IvyGene Cancer Blood Test (targeted NGS; LDT) - Galleri test (targeted NGS, LDT) - PulmoSeek (targeted NGS) - EsoGuard (NGS)	[10,21,27,28,42]
		qMSP (MethyLight)	- Quantitative - Real-time - More sensitive than MSP - Multiplex possibilities	- Disadvantages of bisulfite treatment - Not possible to screen a large number of CpGs - Generation of false-positive and false-negative results - Bias towards the detection of fully methylated templates	- EpiProcolon 2.0 - EpiProLung (BL Reflex Assay) - Cervi-M assay - EarlyTect CRC - Cologuard - Therascreen PITX2 RGQ PCR kit - Real time mS9 CRC - HCCBloodTest - GynTect - QIAsure Methylation test kit - UriFind bladder cancer detection kit - Oral-M - COLVERA™ (LDT) - PredictMDx (LDT) - AssureMDx (LDT) - ConfirmMDx (LDT)	[10,12,21,27,28,32,53]

		Pyrosequencing	- Cost and time-effective - DNA methylation analysis of single gene locus - Quantitative - High throughput capacity - Small PCR amplicons - Quality control for conversion rate	- Disadvantages bisulfite treatment - High cost of instrument - Low detection limit (5%) - Short length of sequence to be analyzed (100 bp)	Therascreen MGMT Pyro kit	[12,21,32]
		ddPCR	- Quantitative - High sensitivity - Analytical sensitivity up to 0.0005% - Low noise - Single molecule DNA template	- Disadvantages of bisulfite treatment - Expensive - Time-consuming optimization	/	[12,17]
		Single cell BS	- Cover high number of CpG sites - Low input through PBAT	- Disadvantages of bisulfite treatment - Information on the DNA strand is lost due to multiple primer extension rounds	/	[15,22,43]
		MIMIC	- Multiplex (17 CpG) - Flexible - Relative low input	- Disadvantages of bisulfite treatment - Expensive	/	[38]
		OBBPA ddPCR	- Advantages ddPCR - More sensitive and specific - Analytical sensitivity of 0.0007%	- Disadvantages of ddPCR	/	[36,37]

	Genome-wide	(cf)RRBS	- Extensive sequencing need is reduced - Quantitative - More cost-effective than WGBS - Increased sample throughput vs WGBS	- Disadvantages of bisulfite treatment - Bias towards high CpG density regions - Less reproducible than WGBS	/	[18,19,39,42]
		Infinium methylation 450K/EPIC (v2.0)	- Relatively low cost - 450K to 850K CpGs - Standardized assay and analysis	- Disadvantages of bisulfite treatment - Not flexible - Large amount of input DNA - Probe-type bias	EPICUP™	[13,17,19,51,52]
		WGBS	- Most powerful - All CpG sites can be studied - Quantitative - Highly reproducible	- Disadvantages of bisulfite treatment - High sequencing cost - Complex computational skills and tools needed (data analysis, storage) - High input DNA	/	[13–15,26,39,42,50]
		ELSA-seq	- High resolution - Low input DNA (500 pg) - Improved library complexity through well-preserved molecular diversity	- Disadvantages of bisulfite treatment - Incomplete duplicate removal due to adaptor tagging - Risk of C→T and G→A artefacts	/	[53,54]
		Hammer-Seq	- Measure methylation of parent and daughter strands simultaneously	- Disadvantages of bisulfite treatment - Large input amount (100 µg DNA) - Time consuming - No distinction in parent-daughter strand	/	[55]

				- High cytosine density regions are overrepresented due to hairpin		
Enzyme-based	*Restriction enzymes*	MSRE- (q)PCR	- Reliable - Simple and cost-effective - qPCR: quantitative	- Time consuming - High input DNA (5µg) (southern blot PCR) - Low throughput and resolution - Potential false-positive results due to incomplete digestion (= inherent to all RE methods)	- Bladder *CARE* (LDT) - EpiCheck	[14,24,27]
		DREAM	- Reproducible - Cost-effective - Quantitative	- Restricted to CCCGGG sites in the genome - High DNA input (1µg) - High-quality DNA only	/	[57,58]
		MED-seq	- Relative low cost - Low DNA input	- Unmethylated CpGs are not detected - Only 50% of genome-wide CpGs is covered	/	[59,60]
		Epi-gSCAR	- Low DNA input - Base pair resolution - Both genomic and epigenomic data	- Need for specific equipment - Expensive - Time consuming (Laborious)	/	[61]
		IMPRESS	- Low DNA input - Low cost - Highly sensitive and specific - Simultaneous detection of mutations and CNVs possible	- Restricted to CpG sites in genome (~40%) - Unmethylated CpGs are not detected - Underexplored	/	
	Other enzymes	TAPS	- High sensitivity - Non-destructive	- Time consuming - Multiple chemical and enzymatic steps that can	/	[42,63,64]

			- High mapping rates and more coverage - Lower sequencing cost	adversely affect conversion efficacy - Difficult individual read mapping		
		Em-Seq	- Locus-specific amplification - High throughput - Low DNA input	- Bias due to incomplete conversion - Underexplored - Expensive	/	[16,64–66]
Affinity-based		MeDIP(-seq)	- Specific antibody to 5mC and 5hmC - Genome-wide coverage - ssDNA: hemi methylation can be verified	- No individual CpG study - Enrichment of CpG dense regions (bias towards hypermethylated regions) - Lower accuracy compared to WGBS/RRBS	/	[13,24,39,42]
		cfMeDIP-seq	- Low input DNA - More cost-effective	- Time consuming - Not yet validated - Difficult to correct for CNVs - No 'unmethylated' signal	/	[13,42,71]
Either of above		Biosensors (optical and electrochemical)	- High sensitivity and specificity - Low cost - Simple and fast - Low DNA input	- Feasibility - Lack of clinical validation - Bisulfite, enzyme or antibody required (with associated disadvantages)	/	[18,65,82]
Direct detection		ONT (Nanopore) TMA-NP	- Native DNA sequencing - Powerful and fast - Simultaneous detection of nucleotide sequence (including mutations and CNVs) and DNA methylation patterns	- High DNA input (1µg) - High error rate (5-20%) - Expensive - No standardized calling pipeline	/	[15,18,22,39,70]

		SMRT (PacBio)	- Native DNA sequencing - Simultaneous detection of nucleotide sequence (including mutations and CNVs) and DNA methylation patterns	- High DNA input (5µg) - High error rate (13-15%) - Expensive - High coverage requirement - Mostly applied in bacteria	/	[15,22,39,75]

DECLARATIONS

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

Isabelle Neefs is supported by a strategic basic PhD fellowship of the Research Foundation Flanders (FWO; IS67523N).

Authors' contributions

IN contributed to the conception of the draft and writing the initial and final draft. **JI** and **MP** contributed to revising the final draft. **GVC** and **KODB** contributed to the conception of the draft and revising the final draft. All authors approved the final version of the draft.

Acknowledgements

Not applicable

REFERENCES

1. Borchiellini M, Ummarino S, Di Ruscio A. The Bright and Dark Side of DNA Methylation: A Matter of Balance. Cells. 2019/10/17. 2019;8.
2. Kulis M, Esteller M. DNA methylation and cancer. Adv Genet. 2010/10/06. 2010;70:27–56.
3. Sestakova S, Krejcik Z, Folta A, Cerovska E, Salek C, Merkerova MD, et al. DNA methylation and hydroxymethylation patterns in acute myeloid leukemia patients with mutations in DNMT3A and IDH1/2 and their combinations. Cancer Biomark. 2019/04/17. 2019;25:43–51.
4. Constancio V, Nunes SP, Henrique R, Jeronimo C. DNA Methylation-Based Testing in Liquid Biopsies as Detection and Prognostic Biomarkers for the Four Major Cancer Types. Cells. 2020/03/11. 2020;9.
5. Ma M, Zhu H, Zhang C, Sun X, Gao X, Chen G. "Liquid biopsy"-ctDNA detection with great potential and challenges. Ann Transl Med. 2015/11/06. 2015;3:235.
6. Han X, Wang JY, Sun YL. Circulating Tumor DNA as Biomarkers for Cancer Detection. Genomics Proteomics Bioinformatics. 2017;15:59–72.
7. José Luis García-Giménez Trygve O. Tollefsbol TU. Epigenetic Biomarkers: New Findings, Perspectives, and Future Directions in Diagnostics. In: García-Giménez JL, editor. Epigenetic Biomarkers and Diagnostics. Academic Press; 2016. p. 680.
8. Dor Y, Cedar H. Principles of DNA methylation and their implications for biology and medicine. Lancet. 2018/08/14. 2018;392:777–86.
9. Joao Lobo Leendert HJ. Looijenga AJMG. Targeted Methylation Analyses: From Bisulfite Treatment to Quantification. In: Amatruda AB and JF, editor. Testicular Germ Cell Tumors: methods and Protocols. Humana Press; 2021.
10. Eggerman T. Multilocus Methylation Assays in Epigenetics. In: Patel V P V, editor. Handbook of Nutrition, Diet, and Epigenetics. Springer, Cham; 2019. p. 2181–202.
11. Pajares MJ, Palanca-Ballester C, Urtasun R, Alemany-Cosme E, Lahoz A, Sandoval J. Methods for analysis of specific DNA methylation status. Methods. 2020/07/09. 2021;187:3–12.
12. Zhao Bapat B. F. The Role of Methylation-Specific PCR and Associated Techniques in Clinical Diagnostics. In: García-Giménez JL, editor. Epigenetic Biomarkers and Diagnosis. Academic Press; 2016. p. 155–73.
13. Huang J, Wang L. Cell-Free DNA Methylation Profiling Analysis-Technologies and Bioinformatics. Cancers (Basel). 2019/11/09. 2019;11.
14. Khodadadi E, Fahmideh L, Khodadadi E, Dao S, Yousefi M, Taghizadeh S, et al. Current Advances in DNA Methylation Analysis Methods. Biomed Res Int. 2021/04/08. 2021;2021:8827516.
15. Li Y. Modern epigenetics methods in biological research. Methods. 2020/07/10. 2021;187:104–13.
16. Vaisvila R, Ponnaluri VKC, Sun Z, Langhorst BW, Saleh L, Guan S, et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. Genome Res. 2021/06/19. 2021;
17. Lissa D, Robles AI. Methylation analyses in liquid biopsy. Transl Lung Cancer Res. 2016/11/09. 2016;5:492–504.
18. Martisova A, Holcakova J, Izadi N, Sebuyoya R, Hrstka R, Bartosik M. DNA Methylation in Solid Tumors: Functions and Methods of Detection. Int J Mol Sci. 2021/05/01. 2021;22.

19. Karilyn E Sant JMG. Methods for Analysis of DNA Methylation. In: Dolinoy SDM and DC, editor. Toxicoepigenetics, Core principles and applications. Academic Press; 2019.
20. Sestakova S, Salek C, Remesova H. DNA Methylation Validation Methods: a Coherent Review with Practical Comparison. Biol Proced Online. 2019/10/05. 2019;21:19.
21. Mikeska T, Craig JM. DNA methylation biomarkers: cancer and beyond. Genes (Basel). 2014/09/18. 2014;5:821–64.
22. Gouil Q, Keniry A. Latest techniques to study DNA methylation. Essays Biochem. 2019/11/23. 2019;63:639–48.
23. Morselli M, Farrell C, Rubbi L, Fehling HL, Henkhaus R, Pellegrini M. Targeted bisulfite sequencing for biomarker discovery. Methods. 2020/08/07. 2021;187:13–27.
24. Tandy L.D. Petrov NCR. The Evolution of New Technologies and Methods in Clinical Epigenetics Research. In: García-Giménez JL, editor. Epigenetic Biomarkers and Diagnostics. Academic Press; 2016. p. 67–89.
25. Masser DR, Stanford DR, Freeman WM. Targeted DNA methylation analysis by next-generation sequencing. J Vis Exp. 2015/03/06. 2015;
26. Kurdyukov S, Bullock M. DNA Methylation Analysis: Choosing the Right Method. Biology (Basel). 2016/01/12. 2016;5.
27. Taryma-Lesniak O, Sokolowska KE, Wojdacz TK. Current status of development of methylation biomarkers for in vitro diagnostic IVD applications. Clin Epigenetics. 2020/07/08. 2020;12:100.
28. Garcia-Gimenez JL, Seco-Cervera M, Tollefsbol TO, Roma-Mateo C, Peiro-Chova L, Lapunzina P, et al. Epigenetic biomarkers: Current strategies and future challenges for their use in the clinical laboratory. Crit Rev Clin Lab Sci. 2017/12/12. 2017;54:529–50.
29. He Q, Chen HY, Bai EQ, Luo YX, Fu RJ, He YS, et al. Development of a multiplex MethyLight assay for the detection of multigene methylation in human colorectal cancer. Cancer Genet Cytogenet. 2010/09/02. 2010;202:1–10.
30. Olkhov-Mitsel E, Zdravic D, Kron K, van der Kwast T, Fleshner N, Bapat B. Novel multiplex MethyLight protocol for detection of DNA methylation in patient tissues and bodily fluids. Sci Rep. 2014/03/22. 2014;4:4432.
31. Singh A, Gupta S, Badarukhiya JA, Sachan M. Detection of aberrant methylation of HOXA9 and HIC1 through multiplex MethyLight assay in serum DNA for the early detection of epithelial ovarian cancer. Int J Cancer. 2020/03/20. 2020;147:1740–52.
32. Beltran-Garcia J, Osca-Verdegal R, Mena-Molla S, Garcia-Gimenez JL. Epigenetic IVD Tests for Personalized Precision Medicine in Cancer. Front Genet. 2019/07/19. 2019;10:621.
33. Florea A-M. Pyrosequencing and Its Application in Epigenetic Clinical Diagnostics. In: García-Giménez JL, editor. Epigenetic Biomarkers and Diagnostics. Academic Press; 2016. p. 175–94.
34. Nyren P. The History of Pyrosequencing((R)). Methods Mol Biol. 2015/06/25. 2015;1315:3–15.
35. Du Y, Yan Z, Song K, Jin J, Xiao L, Sun Z, et al. Development and evaluation of a multiplex droplet digital polymerase chain reaction method for simultaneous detection of five biothreat pathogens. Front Microbiol. 20220728th ed. 2022;13:970973.
36. Menschikowski M, Jandeck C, Friedemann M, Nacke B, Hantsche S, Tiebel O, et al. Identification of rare levels of methylated tumor DNA fragments using an optimized bias based pre-amplification-digital droplet PCR (OBBPA-ddPCR). Oncotarget [Internet]. 2018/12/14. 2018;9:36137–50. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30546833

37. Friedemann M, Horn F, Gutewort K, Tautz L, Jandeck C, Bechmann N, et al. Increased Sensitivity of Detection of RASSF1A and GSTP1 DNA Fragments in Serum of Prostate Cancer Patients: Optimisation of Diagnostics Using OBBPA-ddPCR. Cancers (Basel). 2021/09/11. 2021;13.
38. Schwalbe EC, Hicks D, Rafiee G, Bashton M, Gohlke H, Enshaei A, et al. Minimal methylation classifier (MIMIC): A novel method for derivation and rapid diagnostic detection of disease-associated DNA methylation signatures. Sci Rep. 2017/10/19. 2017;7:13421.
39. Lentini A, Nestor CE. Mapping DNA Methylation in Mammals: The State of the Art. Methods Mol Biol. 2020/08/22. 2021;2198:37–50.
40. Shahryary Y, Hazarika RR, Johannes F. MethylStar: A fast and robust pre-processing pipeline for bulk or single-cell whole-genome bisulfite sequencing data. BMC Genomics. 2020/07/15. 2020;21:479.
41. Liu Z, Wang Z, Jia E, Ouyang T, Pan M, Lu J, et al. Analysis of genome-wide in cell free DNA methylation: progress and prospect. Analyst. 2019/08/23. 2019;144:5912–22.
42. Antonio Galvao GK. Profiling DNA Methylation Genome-Wide in Single Cells. In: Katia Ancelin MB, editor. Epigenetic Reprogramming During Mouse Embryogenesis: Methods and Protocols. Humana Press; 2021. p. 221–40.
43. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 20110802nd ed. 2011;98:288–95.
44. Weichenhan D, Lipka DB, Lutsik P, Goyal A, Plass C. Epigenomic technologies for precision oncology. Semin Cancer Biol. 2020/08/22. 2020;
45. Noguera-Castells A, Garcia-Prieto CA, Alvarez-Errico D, Esteller M. Validation of the new EPIC DNA methylation microarray (900K EPIC v2) for high-throughput profiling of the human DNA methylome. Epigenetics. 2023;18:2185742.
46. Illumina. InfiniumTM MethylationEPIC v2.0 BeadChip. Illumina; 2023.
47. Davalos V, Esteller M. Cancer epigenetics in clinical practice. CA Cancer J Clin. 20221213th ed. 2023;73:376–424.
48. Liang N, Li B, Jia Z, Wang C, Wu P, Zheng T, et al. Ultrasensitive detection of circulating tumour DNA via deep methylation sequencing aided by machine learning. Nat Biomed Eng. 20210615th ed. 2021;5:586–99.
49. Ming X, Zhu B, Zhang Z. Simultaneously measuring the methylation of parent and daughter strands of replicated DNA at the single-molecule level by Hammer-seq. Nat Protoc. 2021/03/10. 2021;16:2131–57.
50. Alonso S, Suzuki K, Yamamoto F, Perucho M. Methylation-Sensitive Amplification Length Polymorphism (MS-AFLP) Microarrays for Epigenetic Analysis of Human Genomes. Methods Mol Biol. 2018/04/02. 2018;1766:137–56.
51. Jelinek J, Madzo J. DREAM: A Simple Method for DNA Methylation Profiling by High-throughput Sequencing. Methods Mol Biol. 2016/09/02. 2016;1465:111–27.

52. Jelinek J, Lee JT, Cesaroni M, Madzo J, Liang S, Lu Y, et al. Digital Restriction Enzyme Analysis of Methylation (DREAM). Methods Mol Biol. 2017/12/11. 2018;1708:247–65.
53. Boers R, Boers J, de Hoon B, Kockx C, Ozgur Z, Molijn A, et al. Genome-wide DNA methylation profiling using the methylation-dependent restriction enzyme LpnPI. Genome Res. 2017/12/10. 2018;28:88–99.

54. Deger T, Boers RG, de Weerd V, Angus L, van der Put MMJ, Boers JB, et al. High-throughput and affordable genome-wide methylation profiling of circulating cell-free DNA by methylated DNA sequencing (MeD-seq) of LpnPI digested fragments. Clin Epigenetics. 2021/10/22. 2021;13:196.
55. Niemoller C, Wehrle J, Riba J, Claus R, Renz N, Rhein J, et al. Bisulfite-free epigenomics and genomics of single cells through methylation-sensitive restriction. Commun Biol. 2021/02/03. 2021;4:153.
56. Neefs I, Vandenhoeck J, Vanpoucke T, Ibrahim J, Suls A, Hoischen A, et al. IMPRESS: Improved Multiplex Methylation Profiling Using Restriction Enzymes and smMIP Sequencing for Highly Sensitive Multi-Cancer Detection. The Journal of Liquid Biopsy. 2023;1:100073.
57. Liu Y, Siejka-Zielinska P, Velikova G, Bi Y, Yuan F, Tomkova M, et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. Nat Biotechnol. 2019/02/26. 2019;37:424–9.
58. Wang T, Loo CE, Kohli RM. Enzymatic approaches for profiling cytosine methylation and hydroxymethylation. Mol Metab. 2021/08/11. 2021;101314.
59. (NEB) NEBI. NEBNext® Enzymatic Methyl-seq Kit [Internet]. 2019. Available from: https://international.neb.com/products/e7120-nebnext-enzymatic-methyl-seq-kit#Product%20Information
60. Sun Z, Vaisvila R, Hussong LM, Yan B, Baum C, Saleh L, et al. Nondestructive enzymatic deamination enables single-molecule long-read amplicon sequencing for the determination of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Genome Res [Internet]. 2021/01/21. 2021; Available from: https://www.ncbi.nlm.nih.gov/pubmed/33468551
61. Ben Maamar M, Sadler-Riggleman I, Beck D, Skinner MK. Genome-Wide Mapping of DNA Methylation 5mC by Methylated DNA Immunoprecipitation (MeDIP)-Sequencing. Methods Mol Biol. 2020/08/22. 2021;2198:301–10.
62. Toska F.J. E; SC. Genome-Wide Techniques for the Study of Clinical Epigenetic Biomarkers. In: García-Giménez JL, editor. Epigenetic Biomarkers and Diagnostics. Academic Press; 2016. p. 119–35.
63. Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature. 2018/11/16. 2018;563:579–83.
64. Pitroda J Dave SK JB. A critical review on Carbon nanotubes. International Journal of Constructive Research in Civil Engineering. 2016;2:36–42.
65. Akbari V, Garant JM, O'Neill K, Pandoh P, Moore R, Marra MA, et al. Megabase-scale methylation phasing using nanopore long reads and NanoMethPhase. Genome Biol. 2021/02/24. 2021;22:68.
66. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. BioMed Central Ltd.; 2020.
67. Liu Y, Rosikiewicz W, Pan Z, Jillette N, Wang P, Taghbalout A, et al. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. Genome Biol. 2021/10/20. 2021;22:295.
68. Y. Suzuki S. Morishita JK. Measuring CpG Methylation by SMRT Sequencing. In: Bredy TW, editor. DNA modifications in the Brain - Neuroepigenetic regulation of Gene expression. Academic Press; 2017. p. 113–25.
69. Roberts CE, Raner GM, Isaacs GD. High Performance Liquid Chromatography Separation of Epigenetic Cytosine Variants. Methods Protoc. 2018;1.

70. Olinski R, Rozalski R, Gackowski D. Mass Spectrometry-Based Analysis of DNA Modifications: Potential Applications in Basic Research and Clinic. Methods Mol Biol. 2020/08/22. 2021;2198:27–35.
71. Dai N, Correa Jr. IR. Liquid Chromatography-Mass Spectrometry Analysis of Cytosine Modifications. Methods Mol Biol. 2020/08/22. 2021;2198:67–78.
72. Boulias K, Greer EL. Detection of DNA Methylation in Genomic DNA by UHPLC-MS/MS. Methods Mol Biol. 2020/08/22. 2021;2198:79–90.
73. Lau BT, Almeda A, Schauer M, McNamara M, Bai X, Meng Q, et al. Single-molecule methylation profiles of cell-free DNA in cancer with nanopore sequencing. Genome Med. 2023;15:1–13.
74. Nazmul Islam M, Yadav S, Hakimul Haque M, Munaz A, Islam F, Al Hossain MS, et al. Optical biosensing strategies for DNA methylation analysis. Biosens Bioelectron. 2016/11/12. 2017;92:668–78

SUPPLEMENTARY MATERIAL

Supplemental Table 1 | Overview of other methylation detection technologies by conversion approach

Approach	Detection	Technology	Advantages	Disadvantages	References
Bisulfite-based		MSP	- High analytical sensitivity (down to 0,1% of methylated alleles can be detected) - Cost-effective	--Disadvantages of bisulfite treatment - Not quantitative - Cannot detect fragmented DNA - Bias towards the detection of fully methylated templates - Low throughput	[1–9]
		MS-SnuPE	- Multiplexing - semi-quantitative - Locus specific	- Disadvantages bisulfite treatment - Limited access to certain CpG's in the PCR product - Radioactivity - Labor-intensive (a.o. primer design, assay)	[8,10–13]
		COBRA	- Easy - Cost-effective	- Disadvantages bisulfite treatment - Limited restriction sites → limited overview DNA methylation - False positives by incomplete digestion - Limited quantification - Underestimation of DNA methylation level - Labor-intensive	[8,11–15]
		MS-DGGE	- Cost-effective - Comprehensive analysis	- Disadvantages bisulfite treatment - Low separation efficiency	[5,11]
		Nested MSP (MN-MSP)	- Simple method for quick validation - Fragmented DNA can be rescued	- Disadvantages bisulfite treatment - Noisy sequencing results - Low sensitivity - Problematic primer design - Limit of 100-300 bp fragments	[16–18]

		MS-MCA	- Easy interpretation in case of equal proportions (un)methylated fragments - Low input (< 10ng)	- Disadvantages bisulfite treatment - Not accurate - Not quantitative - Difficult to multiplex	[11,12,19,20]
		MS-DBA	- Decreased FPR compared to MSP - Simple and cost-effective	- Disadvantages bisulfite treatment - Semi quantitative - Time consuming	[1,21]
		MS-HRM	- Inexpensive and fast - High throughput - Detection limit up to 0,1%	- Disadvantages bisulfite treatment - Not quantitative - Not locus-specific	[3,7,8,22]
		Cold PCR	- 10 to 100 more sensitive than PCR - Cost-effective - Easy to implement - Detection limit up to 0,1%	- Disadvantages bisulfite treatment - Limit of 200 bp fragments	[18,23,24]
		MassARRAY EpiTYPER	- Length of sequence analyzed - Fast and accurate - Highly reproducible - Quantitative - High throughput	- Disadvantages bisulfite treatment - High cost - Low detection limit (5%) - Restricted coverage of CpG's within PCR product - Cleavage products with multiple CpG's are unable to be resolved	[3,8,12,13,17,25,26]
		BSPP	- Flexibility and scalability - Highly specific	- Disadvantages bisulfite treatment	[12,27,28]
		SMRT-BS	- Quantitative - Multiplexing - Relatively low input (50ng)	- Disadvantages of bisulfite treatment - Expensive	[29,30]
		MCTA-seq	- Reduced methylation background - Low input DNA	- Disadvantages of bisulfite treatment - CGIs are enriched - Only CpG tandem regions can be detected	[31–34]
Enzyme-based	*Restriction enzymes*	RFLP/AFLP	- Relatively low cost - Fast - Low DNA input	- Ambiguous interpretation data - Only assess small % of global DNA methylation - Potential false-positive results due to incomplete digestion (= inherent to all RE methods)	[12,18,35–37]

		RLGS	- Quantitative - No prior knowledge needed	- Time consuming - High false-positive rate - Low reproducibility - High quality DNA only	[38–41]
		HELP	- Simple - Positive representation of hypomethylated CpGs	- Not for complex genetic samples - Relative low resolution	[12,42]
		LUMA	- Quantitative - High throughput	- High quality DNA only - Variability with different DNA extraction methods - DNA input can be high (up to 500 ng)	[13,18,43,44]
		CHARM	- Relative low cost - Locus level analysis - Quantitative	- High DNA input - Hybridization-related bias - Relative low resolution	[12,44,45]
		MSCC	- Simple and fast - Accurate	- High quality DNA required	[18,28]
		MS-MLPA	- High throughput screening - Relatively cost-effective - Flexible - Simple data analysis - Fragmented DNA - Low DNA input	- Only semi-quantitative - CpG's limited to HhaI restriction sites - False negatives due to polymorphisms	[1,3,15,46]
		Methyl/MRE-seq	- Cost-effective - Simple and fast	- Lack of whole genome coverage - No absolute methylation levels - No single locus resolution	[2,13,31,47,48]
Affinity-based		ELISA	- Fast and simple protocol - High throughput - Relative low cost	- Relative high input amount (100 ng) - Relative quantification - Cross-reactivity antibodies - Lower sensitivity	[2,18,44,49,50]
		MBD-seq/MethylCap	- Genome wide coverage - Low input DNA - Reproducible	- No individual CpG study - Enrichment of CpG dense regions (bias towards hypermethylated regions) - Areas with less dense 5mC are missed - Lower accuracy compared to WGBS/RRBS - Low recovery of methylated DNA	[12,31,39,51,52]

		MBD-ddPCR	- Highly sensitive - Cost-effective	- Only a limited amount of CpGs - Not validated	[53]
		MIRA-seq	- High sensitivity and specificity - only 1ng DNA input - No denaturation of DNA	- only 100 bp - minimally 2 methylated CpGs are required in the captured fragment	[12,18,54]
Other	*Conversion method*	HPLC-UV	- Quantitative - Fast	- High DNA input (2 µg) - Specialized equipment necessary - Relative high cost/sample	[11,55–57]
		LC/MS-MS	- Sensitive - Quantitative - Relative low input DNA (50 ng) - Not affected by poor quality DNA	- Specialized equipment necessary - Time consuming sample prep - Relative high cost/sample	[13,18,25,58]
		NaIO4/LiBr	- High efficiency - Easy	- Non-selective for sequence - Low sensitivity	[11,59]
		Mwcnts	- No need for chemical/enzymatic modifications - High specificity and precision - Miniaturization - Easy production	- Expensive - Difficult to work with due to size (extremely small) - Might become obsolete due to development of other technologies	[11,60,61]

REFERENCES

1. Zhao Bapat B. F. The Role of Methylation-Specific PCR and Associated Techniques in Clinical Diagnostics. In: García-Giménez JL, editor. Epigenetic Biomarkers and Diagnosis. Academic Press; 2016. p. 155–73.

2. Li Y. Modern epigenetics methods in biological research. Methods. 2020/07/10. 2021;187:104–13.

3. Mikeska T, Craig JM. DNA methylation biomarkers: cancer and beyond. Genes (Basel). 2014/09/18. 2014;5:821–64.

4. Taryma-Lesniak O, Sokolowska KE, Wojdacz TK. Current status of development of methylation biomarkers for in vitro diagnostic IVD applications. Clin Epigenetics. 2020/07/08. 2020;12:100.

5. Aggerholm A, Guldberg P, Hokland M, Hokland P. Extensive intra- and interindividual heterogeneity of p15INK4B methylation in acute myeloid leukemia. Cancer Res. 1999/02/02. 1999;59:436–41.

6. Beltran-Garcia J, Osca-Verdegal R, Mena-Molla S, Garcia-Gimenez JL. Epigenetic IVD Tests for Personalized Precision Medicine in Cancer. Front Genet. 2019/07/19. 2019;10:621.

7. Garcia-Gimenez JL, Seco-Cervera M, Tollefsbol TO, Roma-Mateo C, Peiro-Chova L, Lapunzina P, et al. Epigenetic biomarkers: Current strategies and future challenges for their use in the clinical laboratory. Crit Rev Clin Lab Sci. 2017/12/12. 2017;54:529–50.

8. Mikeska T, Candiloro IL, Dobrovic A. The implications of heterogeneous DNA methylation for the accurate quantification of methylation. Epigenomics. 2010/08/01. 2010;2:561–73.

9. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci U S A. 1996/09/03. 1996;93:9821–6.

10. Gonzalgo ML, Liang G. Methylation-sensitive single-nucleotide primer extension (Ms-SNuPE) for quantitative measurement of DNA methylation. Nat Protoc. 2007/08/19. 2007;2:1931–6.

11. Khodadadi E, Fahmideh L, Khodadadi E, Dao S, Yousefi M, Taghizadeh S, et al. Current Advances in DNA Methylation Analysis Methods. Biomed Res Int. 2021/04/08. 2021;2021:8827516.

12. Yu Liu and Qianjin Lu JL. Laboratory Methods in Epigenetics. In: Qianjin Lu CCC and BCR, editor. Epigenetics and Dermatology. Academic Press; 2015. p. 7–35.

13. Martisova A, Holcakova J, Izadi N, Sebuyoya R, Hrstka R, Bartosik M. DNA Methylation in Solid Tumors: Functions and Methods of Detection. Int J Mol Sci. 2021/05/01. 2021;22.

14. Xiong Z, Laird PW. COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res. 1997/06/15. 1997;25:2532–4.

15. Eggerman T. Multilocus Methylation Assays in Epigenetics. In: Patel V P V, editor. Handbook of Nutrition, Diet, and Epigenetics. Springer, Cham; 2019. p. 2181–202.

16. Palmisano WA, Divine KK, Saccomanno G, Gilliland FD, Baylin SB, Herman JG, et al. Predicting lung cancer by detecting aberrant promoter methylation in sputum. Cancer Res. 2000/11/21. 2000;60:5954–8.

17. Pajares MJ, Palanca-Ballester C, Urtasun R, Alemany-Cosme E, Lahoz A, Sandoval J. Methods for analysis of specific DNA methylation status. Methods. 2020/07/09. 2021;187:3–12.

18. Kurdyukov S, Bullock M. DNA Methylation Analysis: Choosing the Right Method. Biology (Basel). 2016/01/12. 2016;5.

19. Worm J, Aggerholm A, Guldberg P. In-tube DNA methylation profiling by fluorescence melting curve analysis. Clin Chem. 2001/06/28. 2001;47:1183–9.

20. Lorente A, Mueller W, Urdangarin E, Lazcoz P, von Deimling A, Castresana JS. Detection of methylation in promoter sequences by melting curve analysis-based semiquantitative real time PCR. BMC Cancer. 2008/02/27. 2008;8:61.

21. Clement G, Benhattar J. A methylation sensitive dot blot assay (MS-DBA) for the quantitative analysis of DNA methylation in clinical samples. J Clin Pathol. 2005/01/29. 2005;58:155–8.

22. Wojdacz TK, Dobrovic A. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. Nucleic Acids Res. 2007/02/10. 2007;35:e41.

23. Li J, Wang L, Mamon H, Kulke MH, Berbeco R, Makrigiorgos GM. Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. Nat Med. 2008/04/15. 2008;14:579–84.

24. Zuo Z, Jabbar KJ. COLD-PCR: Applications and Advantages. Methods Mol Biol. 2016/02/05. 2016;1392:17–25.

25. Song L, James SR, Kazim L, Karpf AR. Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. Anal Chem. 2005/01/15. 2005;77:504–10.

26. Mikeska T, Bock C, Do H, Dobrovic A. DNA methylation biomarkers in cancer: progress towards clinical implementation. Expert Rev Mol Diagn. 2012/06/19. 2012;12:473–87.

27. Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotechnol. 2009/03/31. 2009;27:353–60.

28. Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol. 2009/03/31. 2009;27:361–8.

29. Yang Y, Scott SA. DNA Methylation Profiling Using Long-Read Single Molecule Real-Time Bisulfite Sequencing (SMRT-BS). Methods Mol Biol. 2017/10/08. 2017;1654:125–34.

30. Yang Y, Sebra R, Pullman BS, Qiao W, Peter I, Desnick RJ, et al. Quantitative and multiplexed DNA methylation analysis using long-read single-molecule real-time bisulfite sequencing (SMRT-BS). BMC Genomics. 2015/05/07. 2015;16:350.

31. Huang J, Wang L. Cell-Free DNA Methylation Profiling Analysis-Technologies and Bioinformatics. Cancers (Basel). 2019/11/09. 2019;11.

32. Liu Z, Wang Z, Jia E, Ouyang T, Pan M, Lu J, et al. Analysis of genome-wide in cell free DNA methylation: progress and prospect. Analyst. 2019/08/23. 2019;144:5912–22.

33. Wen L, Li J, Guo H, Liu X, Zheng S, Zhang D, et al. Genome-scale detection of hypermethylated CpG islands in circulating cell-free DNA of hepatocellular carcinoma patients. Cell Res. 2015/12/02. 2015;25:1376.

34. Zeng H, He B, Yi C, Peng J. Liquid biopsies: DNA methylation analyses in circulating cell-free DNA. J Genet Genomics. 2018/05/01. 2018;45:185–92.

35. Barker D, Schafer M, White R. Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. Cell. 1984/01/01. 1984;36:131–8.

36. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 1995/11/11. 1995;23:4407–14.

37. Alonso S, Suzuki K, Yamamoto F, Perucho M. Methylation-Sensitive Amplification Length Polymorphism (MS-AFLP) Microarrays for Epigenetic Analysis of Human Genomes. Methods Mol Biol. 2018/04/02. 2018;1766:137–56.

38. Hatada I, Hayashizaki Y, Hirotsune S, Komatsubara H, Mukai T. A genomic scanning method for higher organisms using restriction sites as landmarks. Proc Natl Acad Sci U S A. 1991/11/01. 1991;88:9523–7.

39. Toska F.J. E; SC. Genome-Wide Techniques for the Study of Clinical Epigenetic Biomarkers. In: García-Giménez JL, editor. Epigenetic Biomarkers and Diagnostics. Academic Press; 2016. p. 119–35.

40. Rush LJ, Plass C. Restriction landmark genomic scanning for DNA methylation in cancer: past, present, and future applications. Anal Biochem. 2002/08/31. 2002;307:191–201.

41. Spellman P. Gray J. CJ. Cancer Genomics. The Molecular Basis of Cancer. Elsevier Inc; 2008. p. 267–82.

42. Khulan B, Thompson RF, Ye K, Fazzari MJ, Suzuki M, Stasiek E, et al. Comparative isoschizomer profiling of cytosine methylation: the HELP assay. Genome Res. 2006/07/01. 2006;16:1046–55.

43. Karimi M, Johansson S, Stach D, Corcoran M, Grander D, Schalling M, et al. LUMA (LUminometric Methylation Assay)--a high throughput method to the analysis of genomic DNA methylation. Exp Cell Res. 2006/04/21. 2006;312:1989–95.

44. Karilyn E Sant JMG. Methods for Analysis of DNA Methylation. In: Dolinoy SDM and DC, editor. Toxicoepigenetics, Core principles and applications. Academic Press; 2019.

45. Irizarry RA, Ladd-Acosta C, Carvalho B, Wu H, Brandenburg SA, Jeddeloh JA, et al. Comprehensive high-throughput arrays for relative methylation (CHARM). Genome Res [Internet]. 2008/03/05. 2008;18:780–90. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18316654

46. Nygren AO, Ameziane N, Duarte HM, Vijzelaar RN, Waisfisz Q, Hess CJ, et al. Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. Nucleic Acids Res. 2005/08/18. 2005;33:e128.

47. Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, Neff NF, et al. Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. Genome Res. 2009/03/11. 2009;19:1044–56.

48. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 2010/07/09. 2010;466:253–7.

49. Achwal CW, Ganguly P, Chandra HS. Estimation of the amount of 5-methylcytosine in Drosophila melanogaster DNA by amplified ELISA and photoacoustic spectroscopy. EMBO J. 1984/02/01. 1984;3:263–6.

50. Tandy L.D. Petrov NCR. The Evolution of New Technologies and Methods in Clinical Epigenetics Research. In: García-Giménez JL, editor. Epigenetic Biomarkers and Diagnostics. Academic Press; 2016. p. 67–89.

51. Cross SH, Charlton JA, Nan X, Bird AP. Purification of CpG islands using a methylated DNA binding column. Nat Genet. 1994/03/01. 1994;6:236–44.

52. Lissa D, Robles AI. Methylation analyses in liquid biopsy. Transl Lung Cancer Res. 2016/11/09. 2016;5:492–504.

53. Shinjo K, Hara K, Nagae G, Umeda T, Katsushima K, Suzuki M, et al. A novel sensitive detection method for DNA methylation in circulating free DNA of pancreatic cancer. PLoS One. 2020/06/11. 2020;15:e0233782.

54. Rauch T, Pfeifer GP. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. Lab Invest. 2005/07/19. 2005;85:1172–80.

55. Kuo KC, McCune RA, Gehrke CW, Midgett R, Ehrlich M. Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. Nucleic Acids Res. 1980/10/24. 1980;8:4763–76.

56. Roberts CE, Raner GM, Isaacs GD. High Performance Liquid Chromatography Separation of Epigenetic Cytosine Variants. Methods Protoc. 2018;1.

57. Zinellu A, Sotgiu E, Assaretti S, Sotgia S, Paliogiannis P, Pintus G, et al. Evaluation of Global Genomic DNA Methylation in Human Whole Blood by Capillary Electrophoresis UV Detection. J Anal Methods Chem. 2018/01/16. 2017;2017:4065892.

58. Dai N, Correa Jr. IR. Liquid Chromatography-Mass Spectrometry Analysis of Cytosine Modifications. Methods Mol Biol. 2020/08/22. 2021;2198:67–78.

59. Bareyt S, Carell T. Selective detection of 5-methylcytosine sites in DNA. Angew Chem Int Ed Engl. 2007/11/28. 2008;47:181–4.

60. Wang P, Wu H, Dai Z, Zou X. Simultaneous detection of guanine, adenine, thymine and cytosine at choline monolayer supported multiwalled carbon nanotubes film. Biosens Bioelectron. 2011/02/08. 2011;26:3339–45.

61. Pitroda J Dave SK JB. A critical review on Carbon nanotubes. International Journal of Constructive Research in Civil Engineering. 2016;2:36–42.

CHAPTER 2

Colorectal and breast cancer screening: Where are we now?

Isabelle Neefs[1,2], Marc Peeters[2], Guy Van Camp[1,2] and Ken Op de Beeck[1,2]

[1] *Centre of Medical Genetics, University of Antwerp and Antwerp University Hospital, Prins Boudewijnlaan 43, 2650 Edegem, Belgium*

[1] *Centre for Oncological Research Antwerp (CORE), University of Antwerp and Antwerp University Hospital, Universiteitsplein 1, 2610 Wilrijk, Belgium*

[1] *Centre of Medical Genetics, University of Antwerp and Antwerp University Hospital, Prins Boudewijnlaan 43, 2650 Edegem, Belgium*

[2] *Centre for Oncological Research Antwerp (CORE), University of Antwerp and Antwerp University Hospital, Universiteitsplein 1, 2610 Wilrijk, Belgium*

INTRODUCTION

Colorectal cancer and breast cancer

Colorectal cancer (CRC) is the third most common cancer in men and second most common cancer in women worldwide. In both genders, it is the third most deathly cancer type worldwide. In 2020, there were over 1.9 million new cases of CRC worldwide and an estimated number of more than 935,000 deaths [1]. The 5-year survival rate is 91% in early stages but decreases drastically to 14% when CRC is detected in advanced stages. Unfortunately, only 37% of patients are diagnosed in an early stage [2]. Increasing early diagnosis by screening for pre-cancerous lesions and early-stage cancer is crucial to reduce the incidence and mortality, as more effective therapies are available for early stages [3].

Breast cancer (BRCA) is the most common and deadliest cancer type worldwide in women. An estimated number of almost 2.3 million new cases and 685,000 deaths have been reported in 2020 [1]. The 5-year survival rate is 99% for stage I, but decreases to 30% for advanced stages, where still 1 out of 3 patients is diagnosed [6]. As for CRC, increasing early diagnosis of BRCA by screening for in situ carcinomas is essential. In this view, screening programs have been introduced in Europe.

CRC and BRCA detection: current screening methods

Screening and subsequent early detection of cancer is important to decrease the incidence and mortality of this disease. Especially in CRC, research shows that with a two-yearly faecal test, incidence and mortality have been reduced by 18% [9]. Therefore, the CRC population screening program has been established in Flanders since 2013, where people between 50 and 74 years old are invited to take a faecal immunochemical test (FIT). According to the latest data of 2022, the response rate[1] was only 48.3% (45.7% for man, 51.1% for women), which is lower than the EU target participation rate of 65% [14]. Moreover, in 16% of deviating FIT results, patients do not undergo correct follow-up testing. 25.5% of the target population has never even participated in the screening program. Younger people tend to participate less. Although the FIT achieves high specificity (97.1%), the sensitivity, especially for early stages is quite low (67.5% for adenoma). Reported reasons not to participate include mostly psychosocial barriers such as discomfort, fear for a positive result, worrying about false positive results or the detection of CRC, complete ignorance and disgust for the FIT procedure [15–17]. Despite some reasons that are related to factors such as fear, some important reported reasons clearly indicate the need for improvement of the CRC screening program. There are some alternative screening modalities

[1] The response rate is the percentage of people that take up the invitation within 12 months after receiving it

recommended in the EU, such as a flexible sigmoidoscopy (FS) every 5 years or a colonoscopy every 10 years [9]. Despite the major advantage that polyps can be excised during such a screening endoscopy, the bowel preparation needed for these investigations are extensive, and a commonly known barrier for participation [9,16]. Moreover, it has been shown that FIT is more accepted than FS [16]. Other alternative screening modalities are thus needed to increase screening uptake. In this view, few studies have investigated patient preference for blood-based screening tests. In a German study, 97% of patients who refused to undergo a screening colonoscopy, accepted an alternative method. 83% of them selected the methylation *SEPT9* blood test. Only 15% chose a stool test [18]. In a Dutch study, patient preference for either blood-based, stool-based or a combination of both was assessed. Screening was preferred to no screening, with the highest preference for a combination of both blood and stool, followed by blood only [19]. Both studies highlight the need for providing different assays to increase uptake, and for development of novel, blood-based assays.

Even before the CRC screening program, Flanders introduced the BRCA population screening program in 2001. Since then, women between 50-69 years old are invited every two years to participate. Despite the advantages of the program, important limitations persist. For example, in 2022 the response rate was 54.1%, which is significantly lower than the recommended EU guidelines of 70-75% response rates [10,11]. Moreover, 14.4% of the eligible population has never participated in the program and 2.6% of women who do participate, are not followed-up correctly [10,12]. Also, the sensitivity (67.8%) and specificity (98%) of a mammography are not optimal. Reported reasons not to participate are, amongst others, lack of trust in mammographies, worrying about BRCA detection and fear for pain related to the mammography's compression force [13]. Some reasons related to fear will never change with different screening assays, but in view of other reasons, these numbers still highlight the need for introducing alternative screening modalities.

Existing assays and biomarkers for CRC and BRCA detection

CRC

[Adapted from Ferrari A, Neefs I, Hoeck S, Peeters M, Van Hal G, (2021). **Towards Novel Non-Invasive Colorectal Cancer Screening Methods: A Comprehensive Review**. Cancers. https://doi.org/10.3390/cancers 13081820]

Protein Panels

The number of protein molecules indicated by the literature as possible CRC markers in blood is wide. However, only two are currently the main blood-based biomarkers available to detect CRC patients: carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9). CEA is particularly useful when used as a prognostic factor (poor prognostic factor for resectable CRC, cancer progression and recurrence after surgery) but when used as a tool for early detection in a screening setting its sensitivity is low because its levels are strictly related to the tumor stage. Moreover, CEA is not specific for CRC, but a higher level can be caused by liver disease,

pancreatitis, inflammatory bowel disease (IBD), and other malignancies. On the other hand, CA19-9 antigen is even less sensitive and specific for CRC, while it represents a highly reliable marker for the detection of pancreatic and biliary malignancies [20,21].

Combination of Protein and Genes Panels

In 2018, Cohen *et al*. developed CancerSEEK, a blood test that detects multiple types of cancer, including CRC, by combining the detection of circulating free DNA (cfDNA) and protein biomarkers (including CA-125 and CEA) that are released by tumors [22,23]. This test works with an algorithm that weighs the protein and DNA data collected from the blood in order to detect patients who are likely to have a tumor.

Preliminary performance of the test was evaluated in a set of 1005 individuals with known cancers who were compared with 812 healthy controls. Tumor type and location influenced the accuracy of the prediction: the highest accuracy was achieved for colorectal cancer. In particular, specificity of the test was over 99% in eight cancer types: ovarian, liver, stomach, pancreatic, esophageal, colorectal, breast, and lung. Moreover, the tissue of origin was correctly identified in approximately 80% of patients. Although the false-positive rate was low in the trial, it would be expected to be higher in the real-world setting when the test is applied to a healthy population without known diagnosis of cancer. In fact, the authors weighted the results for the actual incidence in the United States and estimated a sensitivity of 55% among the eight cancer types.

In conclusion, the authors stated that multi-analytic tests, such as this one, are not meant to replace other non-blood-based screening tests—such as those for colorectal cancer—but are meant to provide additional information that could help identify and diagnose patients who are at higher risk of having a malignancy [23,24].

Multitarget Stool DNA (MT-sDNA) Test

The American Cancer Society guidelines already recommend the Multitarget Stool DNA test as a feasible option to screen the average-risk population. An example is the FDA-approved MT-sDNA test, called Cologuard®. It is not widely utilized in common practice as other non-invasive tests of the same kind (e.g., FIT) and it is not yet recommended by the official European guidelines as a first line screening test.

Cologuard® is a molecular assay for aberrantly methylated *BMP3* and *NDRG4* promoter regions, mutant *KRAS*, and β-actin, which is used as a reference gene for DNA quantity, combined with an immunochemical assay for human hemoglobin. The study that granted the FDA approval was a cross-sectional study conducted by Imperiale *et al*. It compared the MT-sDNA test with FIT only and the results were generated with the use of a logistic-regression algorithm, with a positive score threshold of 183 or more considered to be positive.

DNA testing showed 92.3% sensitivity for CRC and 42.4% for advanced adenomas. FIT showed 73.8% sensitivity for CRC (p = 0.002) and 23.8% for advanced adenomas (p < 0.001). However, DNA testing showed lower specificity than FIT: 86.6% compared to FIT's 94.9% specificity among patients with nonadvanced or negative findings (p < 0.001) and 89.8% compared to FIT's 96.4%, among those with a negative colonoscopy result (p < 0.001).

These results show that the MT-sDNA test, in asymptomatic persons at average risk for colorectal cancer, detects significantly more cancers than FIT. Therefore, as the authors suggested, being a noninvasive test with a high single-application sensitivity for curable-stage cancer, Cologuard® may provide a suitable option for persons who prefer noninvasive testing, although with lower specificity [25].

ctDNA Methylation—SEPT9 Methylation Assays

One of the genes for which methylation has been linked to CRC development is *SEPT9* [26,27]. Various *SEPT9* gene methylation assays have been developed based on the assumption that the risk of CRC development can be assessed by identifying the degree of DNA methylation of the promoter region of the *SEPT9* gene in peripheral blood.

At present, there is only one *SEPT9* methylation assay kit already approved by the FDA as a valid alternative to CRC screening tests that have already been included in guidelines. This kit, called Epi proColon® 2.0, is to be used on adults, age 50 or above, at average risk for CRC. It uses a real-time polymerase chain reaction (PCR) with a fluorescent hydrolysis probe for the detection of specific methylation in the *SEPT9* DNA target [28,29]. In 2016, the FDA approved the Epi proColon® 2.0, based on the data presented in the PRESEPT study with a 1/3 algorithm [28]. Different studies found a variable sensitivity (ranging between 47–82%) and specificity (ranging between 81–95.9%). They also showed a good sensitivity range for early stage disease (stage I and stage II) around approximately 60% for stage I and 70% for stage II [28,30–32].

The sensitivity of Epi proColon® 2.0 and the commercialized *mSEPT9* assay do not differ from the most widely used Fecal immunochemical tests (FIT) [33–36]. When taking into consideration the fact that FIT has a low compliance in terms of screening uptake [36] and many factors can lead to a false-positive result such as inflammation, infections, ulcers, and hemorrhoids, the *mSEPT9* assay, which is not affected by those factors, could be considered superior, in terms of detection rate, to the fecal test. However, because of the higher cost-effectiveness, FIT remains the first choice among diagnostic tests for CRC screening. It is also notable that different studies found that *mSEPT9* assay sensitivity was further enhanced when it was combined with carcinoembryonic antigen (CEA) or FIT. Consequentially, a combined MS-9 DNA blood test and FIT/CEA may help to achieve a higher detection rate of CRC and may represent a valid option for screening.

BRCA

There are currently no established biomarkers or FDA approved assays for BRCA screening. The only methods that are used and recommended in clinic, are imaging technologies such as the mammography, ultrasound etc [37]. Nevertheless, there are some important molecular markers in use for diagnosing BRCA, as it is a very heterogeneous disease with various molecular aberrations. As such, molecular subtyping per case is of extreme importance in BRCA diagnostics, mostly for treatment and follow-up purposes [38].

Immunohistochemical (IHC) analysis for diagnostic purposes

The European Society for Medical Oncology (ESMO) guidelines for BRCA recommend IHC to assess classical molecular markers such as the estrogen receptor (ER), progesterone receptor (PR) and oncogenic human epidermal growth factor receptor 2 (HER2). Protein staining of these markers on tumor tissues is nowadays used as gold standard methods for subtyping BRCA [38,39]. Apart from histological subtyping of BRCA into lobular and ductal carcinomas [37], there are four main groups into which BRCA can be divided based on the IHC staining: 1) Hormone receptor (HR) positive, HER2 negative patients, 2) HR positive, HER2 positive patients, 3) HR negative, HER2 positive patients and 4) HR negative, HER2 negative patients. The latter are known as triple negative breast cancers (TNBC). Knowing the subtype is mostly important for treatment options, as there exist anti-hormonal and anti-HER2 therapies [38]. Furthermore, it is known that, for example HER2 positive carcinomas are aggressive and lead to a poor prognosis [37]. Beside receptors, also the Ki-67 proliferation index is used as a prognostic biomarker in BRCA [40,41].

Serological markers for diagnostic purposes

There are very few serological markers described that are used in routine diagnostics. The previously mentioned Ki-67 is traditionally assessed in tumor tissue, but has potential as a serological biomarker. Furthermore, CA 15-3 is used in diagnosis of BRCA. It is a protein component of MUC 1, which is found in epithelial cells. Besides its use in BRCA detection, it is also frequently used in monitoring BRCA. However, CA 15-3 levels can also be elevated in other conditions including gastrointestinal and lung neoplasm, which makes it a suboptimal biomarker [41].

Genetic alterations for diagnostic purposes

There are a few genetic biomarkers and assays testing for them that are used in routine diagnostics. First of all, the PAM50 assay analyses the gene expression of 50 genes. Based on the mRNA expression that is measured with the RT-qPCR assay, BRCA can be classified into four 'intrinsic' or molecular subtypes: 1) Luminal A, 2) Luminal B, 3) HER2 enriched and 4) Basal-like BRCA. Knowing these subtypes enables tailoring the treatment of the patient [37,38]. Second, there exist multi-gene panels for screening high-risk patients with BRCA

predisposition. The best-known and most significant high-risk genes for predisposition to BRCA, are *BRCA1* and *BRCA2* [42]. Genetic counselling is used to follow-up these high-risk patients, using e.g. the TruRisk panel. This assay consists of different genes, including the *BRCA1/2, CHEK2, RAD51C, PTEN* and *TP53* genes [37,42]. Another routinely used assay is the FDA-approved MammaPrint assay, which was validated for use in BRCA through the MINDACT trial [43]. With the MammaPrint, 70 genes are evaluated to determine the risk of distant metastasis in early stage breast cancer. The outcome is mostly interesting to inform decisions regarding therapy [43]. Lastly, the Oncotype DX assay is another option to analyse expression of multiple genes (n=21) in tumor tissue samples [41].

Beside the assays, there are some important genes that are routinely tested for in the clinic. These include the *PIK3CA, GATA3, MAP3K1, ATM, PALB2, BRIP1* and *MCM4* genes [39,41]. All genes that are evaluated, are mostly important for therapy initiation decisions and follow-up, and to a lesser extent for detection of BRCA. Importantly, genes and proteins discussed are only in use for Luminal and HER2 carcinomas. For TNBC, biomarkers are currently being developed but not yet used in clinical diagnostics [41].

Despite all advances, there is still room to improve screening of BRCA and CRC by using other biomarker types. Our research is tailored towards DNA methylation, as it has been demonstrated extensively that the methylome is a powerful source for biomarkers, and needs further investigation.

DNA methylation and its role in CRC and BRCA

Epigenetics plays a crucial role in cancer development. One of the most studied epigenetic alterations is DNA methylation. It occurs mostly on the fifth position of the cytosine of a CpG dinucleotide in so-called CpG islands, which are regions in the genome with a high density of CpG dinucleotides [44]. In healthy cells, DNA methylation ensures activity of required genes via hypomethylation of promotor regions of tumor suppressor genes as well as genome stability by hypermethylation of the gene body and repetitive elements [45]. In cancer, the methylation patterns are altered. These changes occur early in cancer development and typically silence the expression of tumor suppressor genes by hypermethylation of the promotor region. In addition, the gene body and repetitive elements are hypomethylated [46].

In oncology, mutations are the most researched biomarkers. However, based on the involvement of methylation in carcinogenesis, methylation patterns are now also considered as potential biomarkers. It has been demonstrated for both CRC and BRCA that epigenetic alterations found in the primary tumor can also be detected in ctDNA extracted from plasma, underlining the potential of methylated ctDNA (MetctDNA) as a biomarker [47,48]. The use of DNA methylation biomarkers has some major advantages over using mutations. Methylation occurs very early, possibly before actual neoplastic transformation in cancer development, which renders it especially interesting for screening and early diagnosis. Given that no prior knowledge is needed on

the tumor molecular profile, methylation biomarkers are more universal than mutation markers. Since methylation-based tests can be used off the shelf, they are much faster and cheaper to use [49,50].

For both CRC and BRCA, several genes have already been examined for their biomarker potential in early diagnosis. Examples include methylation of *APC* for BRCA and *SEPT9* for CRC [48,51]. *SEPT9* promotor methylation for the detection of CRC has been commercialized as the Epi proColon® test (Epigenomics AG). It was the first blood-based screening test to be approved by the FDA in 2016. Its sensitivity ranges from 42 - 96% and its specificity ranges from 79 - 99% in different studies performed by separate research groups [51]. More research is needed to improve the overall sensitivity of the test. Many more biomarkers have been described, but these merely consider a single cancer type or gene of interest. Simultaneously analyzing several genes or regions of interest could improve the overall sensitivity of biomarker tests drastically.

Minimally-invasive cancer detection using liquid biopsies

At present, tumor tissue sampling remains the gold standard for detection and diagnosis of any cancer. Due to several difficulties that arise when taking solid biopsies, liquid biopsies are a recent and popular alternative in oncology research. Liquid biopsies are defined as the assessment of circulating components in bodily fluids. It has already been demonstrated that alterations found in a tumor tissue sample can be found by surrogate markers in liquid biopsies [47,49]. Important advantages of liquid biopsies include minimal invasiveness, the ability to represent the whole (epi)genetic landscape of the tumor including that of metastases and its accessibility, allowing repeated sampling and close follow-up [46,49,52]. One of the most investigated surrogate markers is circulating tumor DNA (ctDNA), which is the tumor-specific fraction of all circulating cell-free DNA (cfDNA). It is mostly shed passively into the bloodstream via apoptosis or necrosis of the tumor cells and a fraction also is excreted actively via extra cellular vesicles [53]. The amounts of ctDNA shed into the bloodstream vary from undetectable (<0.1%) up to >50% of total cfDNA [47]. As all tumoral regions are expected to release ctDNA, it reflects the cumulative tumor burden. Previous research has shown that both BRCA patients and CRC patients shed ctDNA into the bloodstream, even in early stages [47,48]. Lastly, studies indicate that current screening methods are considered unpleasant and therefore, blood-based testing could be a solution to increase screening uptake [54–57].

The hurdles in methylated (ct)DNA analysis

Nowadays, bisulfite sequencing is considered the gold standard for DNA methylation analysis. One of the major disadvantages of bisulfite sequencing is the significant amount of DNA degradation, which is especially important when using fragmented samples with low DNA concentration such as ctDNA. Alternatively, bisulfite-free techniques including methylation-sensitive restriction enzymes (MSRE) are available. All technologies with

their (dis)advantages were already described in more detail in **chapter 1**. Besides the problems inherent to the methylation detection techniques, another problem arises when detecting ctDNA. As mentioned before, ctDNA is fragmented and its concentration in blood is very low [44]. Consequently, only a fraction of the genome could be represented in liquid biopsies. Detection of methylation in one gene may generate false-negative results since the fraction may not contain the specific gene of interest, limiting the sensitivity. However, this limitation can be overcome by simultaneous detection of the methylation status of multiple genes or regions in a multiplex assay. Moreover, this increases the sensitivity of the test, since the chance of missing all genes or regions is significantly lower. Unfortunately, up to this moment, there is no cost-effective technique that allows multi-region methylation analysis. In conclusion, there is an urgent need for a bisulfite-free detection method for the simultaneous analysis of multiple methylated regions.

REFERENCES

1. World Health Organisation and International Agency for Research on Cancer. GLOBOCAN - Cancer Today. 2020.

2. Colorectal Cancer: Statistics | Cancer.Net [Internet]. [cited 2024 Jan 25]. Available from: https://www.cancer.net/cancer-types/colorectal-cancer/statistics

3. Labianca R, Nordlinger B, Beretta GD, Mosconi S, Mandalà M, Cervantes A, et al. Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2013;24:vi64–72.

4. Belgisch Kanker Register. Jaar rapport - Bevolkingsonderzoek naar kanker 2022. 2022.

5. Issa IA, NouredDine M. Colorectal cancer screening: An updated review of the available options. World J Gastroenterol [Internet]. 2017 [cited 2024 Jan 25];23:5086–96. Available from: https://pubmed.ncbi.nlm.nih.gov/28811705/

6. Breast Cancer: Statistics | Cancer.Net [Internet]. [cited 2024 Jan 25]. Available from: https://www.cancer.net/cancer-types/breast-cancer/statistics

7. Von Euler-Chelpin M, Lillholm M, Vejborg I, Nielsen M, Lynge E. Sensitivity of screening mammography by density and texture: A cohort study from a population-based screening program in Denmark. Breast Cancer Research [Internet]. 2019 [cited 2024 Jan 25];21:1–7. Available from: https://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-019-1203-3

8. Niell BL, Freer PE, Weinfurtner RJ, Arleo EK, Drukteinis JS. Screening for Breast Cancer. [cited 2024 Jan 25]; Available from: http://dx.doi.org/10.1016/j.rcl.2017.06.004

9. Marcellinaro R, Spoletini D, Grieco M, Avella P, Cappuccio M, Troiano R, et al. Colorectal Cancer: Current Updates and Future Perspectives. J Clin Med. 2023;13:40.

10. Centrum voor Kankeropsporing (CvKO) BKR (BCR). Jaarfiche 2023 [Internet]. Available from: https://borstkanker.bevolkingsonderzoek.be/nl/literatuur-bk.

11. European Commission. Europe's Beating Cancer Plan Communication from the commission to the European Parliament and the Council.

12. Centrum voor Kankeropsporing (CvKO) BKR (BCR). Jaarrapport 2022.

13. Sterlingova T, Nylander E, Almqvist L, Møller Christensen B. Factors affecting women's participation in mammography screening in Nordic countries: A systematic review. Radiography. 2023;29:878–85.

14. Joint Statement Call for action on Colorectal Cancer Screening in the EU.

15. Le Bonniec A, Meade O, Fredrix M, Morrissey E, O'Carroll RE, Murphy PJ, et al. Exploring non-participation in colorectal cancer screening: A systematic review of qualitative studies. Soc Sci Med. 2023;329:116022.

16. Ali O, Gupta S, Brain K, Lifford KJ, Paranjothy S, Dolwani S. Acceptability of alternative technologies compared with faecal immunochemical test and/or colonoscopy in colorectal cancer screening: A systematic review. J Med Screen. 2023;30:14–27.

17. Berg-Beckhoff G, Leppin A, Nielsen JB. Reasons for participation and non-participation in colorectal cancer screening. Public Health. 2022;205:83–9.

18. Adler A, Geiger S, Keil A, Bias H, Schatz P, Devos T, et al. Improving compliance to colorectal cancer screening using blood and stool based tests in patients refusing screening colonoscopy in Germany. 2014 [cited 2024 Mar 7]; Available from: http://www.biomedcentral.com/1471-230X/14/183

19. Benning TM, Dellaert BGC, Dirksen CD, Severens JL. Preferences for potential innovations in non-invasive colorectal cancer screening: A labeled discrete choice experiment for a Dutch screening campaign. 2014 [cited 2024 Mar 7]; Available from: https://www.tandfonline.com/action/journalInformation?journalCode=ionc20

20. Alves Martins BA, de Bulhões GF, Cavalcanti IN, Martins MM, de Oliveira PG, Martins AMA. Biomarkers in Colorectal Cancer: The Role of Translational Proteomics Research. Front Oncol [Internet]. 2019 [cited 2024 Mar 12];9:476229. Available from: www.frontiersin.org

21. Gonzalez-Pons M, Cruz-Correa M. Colorectal Cancer Biomarkers: Where Are We Now? Biomed Res Int. 2015;2015.

22. De Rubis G, Rajeev Krishnan S, Bebawy M. Liquid Biopsies in Cancer Diagnosis, Monitoring, and Prognosis. Trends Pharmacol Sci. 2019;40:172–86.

23. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science (1979) [Internet]. 2018 [cited 2024 Mar 12];359:926–30. Available from: https://www.science.org/doi/10.1126/science.aar3247

24. Advancing Cancer Screening with Liquid Biopsies. Cancer Discov [Internet]. 2018 [cited 2024 Mar 12];8:256. Available from: /cancerdiscovery/article/8/3/256/6185/Advancing-Cancer-Screening-with-Liquid

25. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget Stool DNA Testing for Colorectal-Cancer Screening. New England Journal of Medicine [Internet]. 2014 [cited 2024 Mar 12];370:1287–97. Available from: https://www.nejm.org/doi/10.1056/NEJMoa1311194

26. Hall PA, Russell SEH. The pathobiology of the septin gene family. J Pathol [Internet]. 2004 [cited 2024 Mar 12];204:489–505. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/path.1654

27. Tóth K, Galamb O, Spisák S, Wichmann B, Sipos F, Valcz G, et al. The influence of methylated septin 9 gene on RNA and protein level in colorectal cancer. Pathology and Oncology Research [Internet]. 2011 [cited 2024 Mar 12];17:503–9. Available from: https://link.springer.com/article/10.1007/s12253-010-9338-7

28. Song L, Jia J, Peng X, Xiao W, Li Y. The performance of the SEPT9 gene methylation assay and a comparison with other CRC screening tests: A meta-analysis. Scientific Reports 2017 7:1 [Internet]. 2017 [cited 2024 Mar 12];7:1–12. Available from: https://www.nature.com/articles/s41598-017-03321-8

29. Premarket Approval (PMA).

30. Wu D, Zhou G, Jin P, Zhu J, Li S, Wu Q, et al. Detection of Colorectal Cancer Using a Simplified SEPT9 Gene Methylation Assay Is a Reliable Method for Opportunistic Screening. J Mol Diagn [Internet]. 2016 [cited 2024 Mar 12];18:535–45. Available from: https://pubmed.ncbi.nlm.nih.gov/27133379/

31. Chen CH, Yan SL, Yang TH, Chen SF, Yeh YH, Ou JJ, et al. The Relationship between the Methylated Septin-9 DNA Blood Test and Stool Occult Blood Test for Diagnosing Colorectal Cancer in Taiwanese People. J Clin Lab Anal [Internet]. 2017 [cited 2024 Mar 12];31:e22013. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/jcla.22013

32. Xie L, Jiang X, Li Q, Sun Z, Quan W, Duan Y, et al. Diagnostic value of methylated Septin9 for colorectal cancer detection. Front Oncol [Internet]. 2018 [cited 2024 Mar 12];8:388984. Available from: www.frontiersin.org

33. Tóth K, Sipos F, Kalmár A, Patai Á V., Wichmann B, Stoehr R, et al. Detection of methylated SEPT9 in plasma is a reliable screening method for both left- and right-sided colon cancers. PLoS One [Internet]. 2012 [cited 2024 Mar 12];7. Available from: https://pubmed.ncbi.nlm.nih.gov/23049919/

34. Whitlock EP, Lin JS, Liles E, Beil TL, Fu R. Screening for colorectal cancer: a targeted, updated systematic review for the U.S. Preventive Services Task Force. Ann Intern Med [Internet]. 2008 [cited 2024 Mar 12];149:638–58. Available from: https://pubmed.ncbi.nlm.nih.gov/18838718/

35. LeeVan E, Pinsky P. Predictive Performance of Cell-Free Nucleic Acid-Based Multi-Cancer Early Detection Tests: A Systematic Review. Clin Chem [Internet]. 2024 [cited 2024 Feb 1];70:90–101. Available from: https://dx.doi.org/10.1093/clinchem/hvad134

36. The Lancet Gastroenterology & Hepatology. Improving uptake of colorectal cancer screening. Lancet Gastroenterol Hepatol. 2017;2:767.

37. Bonetti G, Madeo G, Michelini S, Ricci M, Cestari M, Gadler M, et al. Omics sciences and precision medicine in breast and ovarian cancer. Clin Ter. 2023;174:104–18.

38. Keup C, Kimmig R, Kasimir-Bauer S. The Diversity of Liquid Biopsies and Their Potential in Breast Cancer Management. Cancers (Basel) [Internet]. 2023 [cited 2024 Feb 1];15. Available from: /pmc/articles/PMC10670968/

39. Pankotai-Bodó G, Oláh-Németh O, Sükösd F, Pankotai T. Routine molecular applications and recent advances in breast cancer diagnostics. J Biotechnol. 2024;380:20–8.

40. Herrera Juarez M, Tolosa Ortega P, Sanchez de Torre A, Ciruelos Gil E. Biology of the Triple-Negative Breast Cancer: Immunohistochemical, RNA, and DNA Features. Breast Care. 2020;15:208–16.

41. Lopez-Gonzalez L, Sanchez Cendra A, Sanchez Cendra C, Roberts Cervantes ED, Espinosa JC, Pekarek T, et al. Exploring Biomarkers in Breast Cancer: Hallmarks of Diagnosis, Treatment, and Follow-Up in Clinical Practice. Medicina (B Aires). 2024;60:168.

42. Speiser D, Bick U. Breast Care Primary Prevention and Early Detection of Hereditary Breast Cancer. Breast Care [Internet]. 2023 [cited 2024 Feb 19];18:450–6. Available from: http://www.karger.com/Services/

43. William Audeh LBHKHTGS. View of Prospective Validation of a Genomic Assay in Breast Cancer: The 70-gene MammaPrint Assay and the MINDACT Trial [Internet]. [cited 2024 Mar 12]. Available from: https://www.ama.ba/index.php/ama/article/view/355/pdf

44. Elazezy M, Joosse SA. Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management. Comput Struct Biotechnol J [Internet]. 2018 [cited 2024 Jan 25];16:370. Available from: /pmc/articles/PMC6197739/

45. Kulis M, Esteller M. DNA methylation and cancer. Adv Genet [Internet]. 2010/10/06. 2010;70:27–56. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20920744

46. Warton K, Mahon KL, Samimi G. Methylated circulating tumor DNA in blood: power in cancer prognosis and response. Endocr Relat Cancer. 2016;23:R157–71.

47. Moati E, Taly V, Didelot A, Perkins G, Blons H, Taieb J, et al. Role of circulating tumor DNA in the management of patients with colorectal cancer. Clin Res Hepatol Gastroenterol. 2018;42:396–402.

48. Tzanikou E, Lianidou E. The potential of ctDNA analysis in breast cancer. Crit Rev Clin Lab Sci [Internet]. 2020 [cited 2024 Jan 25];57:54–72. Available from: https://www.tandfonline.com/doi/abs/10.1080/10408363.2019.1670615

49. Han X, Wang JY, Sun YL. Circulating Tumor DNA as Biomarkers for Cancer Detection. Genomics Proteomics Bioinformatics [Internet]. 2017;15:59–72. Available from: <Go to ISI>://WOS:000401309200002

50. Ma M, Zhu H, Zhang C, Sun X, Gao X, Chen G. "Liquid biopsy"-ctDNA detection with great potential and challenges. Ann Transl Med [Internet]. 2015/11/06. 2015;3:235. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26539452

51. Song L, Jia J, Peng X, Xiao W, Li Y. The performance of the SEPT9 gene methylation assay and a comparison with other CRC screening tests: A meta-analysis. Sci Rep [Internet]. 2017 [cited 2024 Jan 25];7:3032–3032. Available from: https://europepmc.org/articles/PMC5465203

52. Vymetalkova V, Cervena K, Bartu L, Vodicka P. Circulating Cell-Free DNA and Colorectal Cancer: A Systematic Review. Int J Mol Sci. 2018;19:3356.

53. Stejskal P, Goodarzi H, Srovnal J, Hajdúch M, Van 't Veer LJ, Jesus M, et al. Circulating tumor nucleic acids: biology, release mechanisms, and clinical relevance. Mol Cancer [Internet]. 2023 [cited 2024 Jan 30];22:15. Available from: http://creativecommons.org/licenses/by/4.0/.TheCreativeCommonsPublicDomainDedicationwaiver

54. Osborne JM, Wilson C, Moore V, Gregory T, Flight I, Young GP. Sample preference for colorectal cancer screening tests: Blood or stool? Open J Prev Med [Internet]. 2012 [cited 2024 Jan 30];2:326–31. Available from: http://dx.doi.org/10.4236/ojpm.2012.23047

55. Gelhorn H, Ross MM, Kansal AR, Fung ET, Seiden M V., Krucien N, et al. Patient Preferences for Multi-Cancer Early Detection (MCED) Screening Tests. Patient [Internet]. 2023 [cited 2024 Jan 30];16:43–56. Available from: https://link.springer.com/article/10.1007/s40271-022-00589-5

56. Berg-Beckhoff G, Leppin A, Nielsen JB. Reasons for participation and non-participation in colorectal cancer screening. Public Health. 2022;205:83–9.

57. Le Bonniec A, Meade O, Fredrix M, Morrissey E, O'Carroll RE, Murphy PJ, et al. Exploring non-participation in colorectal cancer screening: A systematic review of qualitative studies. Soc Sci Med. Elsevier Ltd; 2023.

Aims | Outline

AIMS AND OUTLINE OF THE PHD THESIS

In 2019, our research group demonstrated the potential of the *GSDME* gene as a universal methylation marker for differential detection of 14 types of cancer [23]. This study was later expanded to multi-cancer genome-wide methylation profiling with a focus on the 14 most prevalent cancers, including BRCA and CRC [24]. Due to the large number of CpGs in the human genome, the ones that can be used in a biomarker setting were carefully selected based on having the largest differences in methylation between normal and tumor tissues. Consistently differentially methylated CpG sites (DMCs) in all studied cancer types were identified, in view of developing an assay for multi-cancer detection. For example, a total of 77,302 DMCs corresponding to 12,350 different genes common for breast, CRC and lung cancer with calculated AUCs ranging from 0.84-0.99 were found [24]. This demonstrates that CpG methylation can reliably be used as both a detection (presence or absence) and differential (tumor type) biomarker for cancer. The prediction models were not influenced by tumor stage or patient age. Importantly, the large number of identified DMCs highlighted the need for a novel highly multiplexable targeted methylation detection method. The development of this method was a major part in this PhD thesis (see **aim 2** and **chapter 6**).

DNA methylation biomarkers are gaining attention in the oncological research field, but current detection methods leave much to be desired in view of sensitivity and cost. Therefore, in this PhD research, we have explored the epigenetic landscape of cancers, in particular colorectal cancer and breast cancer, with the general aims of identifying new DNA methylation cancer biomarkers and detecting them in tissue biopsies using newly developed assays.

The following aims were formulated for this PhD thesis:

Aim I. To explore conventional and new biomarkers, with a focus on CRC and DNA methylation

In literature, there is little to no information concerning clinicopathological characteristics and molecular alterations in CRC interval cancers. Therefore, we investigated these features in stage IV screen-detected and FIT-interval CRCs. We focused on the currently reported, thus classical biomarkers. The results are described in **chapter 3**. With regard to previous epigenomic findings of our own, and the lack of comparison in the methylation levels between normal, adenoma and carcinoma tissue in CRC, we aimed to explored DNA methylation as a biomarker for early CRC detection. We aimed to investigate whether there is differential methylation between normal colorectal tissue, precancerous lesions (adenomas) and invasive carcinomas to unravel potential biomarkers that can discriminate between the three tissue types. The results are described in **chapter 4.**

Aim II. To create novel assays using both classic and novel technologies for the analysis of (ct)DNA methylation biomarkers

As outlined in the introduction, DNA methylation detection is not optimal yet. Aiming to tackle this problem, one of the goals was to develop novel assays using multiplexing strategies. In one study, ddPCR was chosen as it is a well-known method, but only gained attention for DNA methylation detection recently. Moreover, ddPCR assays can be easily translated towards the clinic. However, only a limited number of targets can be investigated. This is also highlighted in the results of **chapter 5.** As there was no ideal method present, one of the objectives was to focus on the development of a new technology, termed IMPRESS, for the analysis of thousands of methylated CpGs simultaneously. This technology and all achieved results are described in detail in **chapter 6**.

Aim III. To validate our newly developed DNA methylation biomarker assays

As outlined in the rationale, screening for CRC and BRCA is not optimal yet. Therefore, the last goal of this thesis was to development a specific IMPRESS assay for CRC and BRCA. To achieve this goal, previously found biomarkers from **chapter 4** and the technology from **chapter 6** are used in combination to develop the assays. The aim was to validate the new assays in both fresh frozen tissue and liquid biopsies.

PART II | Exploring novel biomarkers

CHAPTER 3

Clinicopathological and molecular differences between stage IV screen-detected and interval colorectal cancers in the Flemish screening program

(Revised manuscript submitted in *Frontiers in Oncology*)

Isabelle Neefs[1,2], Thuy Ngan Tran[3,4], Allegra Ferrari[3,5], Sharon Janssens[6], Koen Van Herck[6], Ken Op de Beeck[1,2], Guy Van Camp[1,2], Marc Peeters[2], Erik Fransen[1], Sarah Hoeck[3,4] and Guido Van Hal[3,4]

1 Centre of Medical Genetics, University of Antwerp and Antwerp University Hospital, Prins Boudewijnlaan 43, 2650 Edegem, Belgium

2 Centre for Oncological Research Antwerp (CORE), University of Antwerp and Antwerp University Hospital, Universiteitsplein 1, 2610 Wilrijk, Belgium

3 Research group on Social Epidemiology and Health Policy, Department of Family Medicine and Population Health (FAMPOP), University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium

4 Centre for Cancer Detection, Ruddershove 4, 8000 Bruges, Belgium

5 Department of Health Sciences (DISSAL), Via Balbi 5, University of Genoa, 16123 Genoa, Italy

6 Belgian Cancer Registry, Koningsstraat 215, 1210 Brussels, Belgium

ABSTRACT

Introduction: Interval cancer (IC) is an important quality indicator in colorectal cancer (CRC) screening. Previously, we found that fecal immunochemical test (FIT) ICs are more common in women, older age, right-sided tumors, and advanced stage. Here, we extended our existing stage IV patient cohort with clinicopathological and molecular characteristics, to identify factors associated with FIT-IC.

Methods: Logistic regression models were fit to identify variables associated with the odds of having a stage IV FIT-IC. Multivariate models were corrected for gender, age, and location.

Results: A total of 292 screen-detected (SD) CRCs and 215 FIT-IC CRCs were included. FIT-IC CRC had 5 fold higher odds to be a neuroendocrine tumor (NET) and 2.5 fold higher odds to have lymphovascular invasion. Interestingly, some variables lost significance upon accounting for location. Thus, tumor location is a critical covariate that should always be included when evaluating factors related to FIT-IC.

Conclusions: We identified NETs and lymphovascular invasion as factors associated with increased odds of having a stage IV FIT-IC. Moreover, we highlight the importance of tumor location as a covariate in evaluating FIT-IC related factors. More research across all stages is needed to clarify how these insights might help to optimize the Flemish CRC screening program.

INTRODUCTION

Colorectal cancer (CRC) is one of the leading causes of cancer-related deaths worldwide. In Belgium, CRC is the third most common cancer in men and second most common in women [1]. In 2021, almost 8 000 persons were newly diagnosed with CRC in Belgium. In Flanders (58% of the Belgian population), the age-standardized incidence rates (world standard population) for CRC were 33.9 per 100,000 person-years for men and 24.2 per 100,000 person-years for women in 2021 [2].

Most CRC tumors slowly develop through multiple steps starting from precancerous lesions. Over time, both morphologic, histologic and also molecular modifications can accumulate, leading to invasive tumors. As for most cancer types, prognosis of CRC depends on the tumour stage at diagnosis, and can be drastically improved by early detection [3]. This is clearly shown by the 5-year survival rates in Flanders, which are 96% for stage I and only 21% for stage IV in persons between 50 through 74 years old [4]. Moreover, in high-income countries, incidence rates of CRC have been decreasing as a result of the implementation of screening programs. If these screening programs are effective and in place in all European countries, an additional 80 000 CRC deaths could be prevented yearly [5]. It is clear that screening of CRC is an important tool to reduce incidence and mortality. In Flanders, CRC screening was introduced in October 2013 for all persons aged 56 through 74 years. The starting age was gradually lowered from 56 in 2013 to 50 in 2020. Currently, the program offers a biennial, free of charge fecal immunochemical test (FIT: FOB Gold, Sentinel Diagnostics, Milan, Italy) to individuals from 50 through 74 years old [6]. More details about the program and the inclusion/exclusion criteria for screening have previously been described [7].

One of the important quality indicators to consider in the CRC screening program is the occurrence of FIT interval cancers (FIT-ICs). FIT-IC are defined as CRC diagnosed after a negative FIT, but before the next FIT invitation (24 months). On the other hand, screen-detected (SD) cancers are defined as a CRC diagnosed after a positive FIT, within 6 months after the first follow-up colonoscopy and before the next recommended FIT invitation (24 months) [7,8]. Previous research by Tran *et al.* [7] about FIT-ICs already showed an overall FIT-IC proportion of 13%. Importantly, they also reported a significantly higher risk of having a FIT-IC versus an SD CRC for female gender, older age, right-sided location, high differentiation grade and stage IV compared to stage I participants in the in-study population, for which the latter showed the strongest association with the risk of having FIT-IC (OR= 7.15 [5.76 – 8.88]) When looking at stage IV alone, the FIT-IC proportion was 45% [7].

In addition to these findings, it has been described that relative 5-year survival rates for FIT-IC are drastically lower compared to SD-ICs (67% vs 94% respectively). The larger proportion of stage IV cancers in the FIT-IC group lead to this lower relative survival chances [9]. Furthermore, FIT still has room for improvement in

sensitivity, which is currently 79% (ranging from 70% – 86% in different studies) for stage IV [9]. Specific and reliable markers for ICs can help here to increase the SD IV proportion.

Unfortunately, literature about clinicopathological and molecular features of FIT-IC cancers is rare, and large studies are still lacking. Therefore, we wanted to extend the previous study by Tran *et al.* [7], by exploring clinicopathological and molecular characteristics of FIT-IC versus FIT-SD CRC. Taking into account the results of Tran *et al.,* suggesting that stage IV is the most interesting subgroup for further investigations, we expanded the existing data of all advanced stage (IV) patients with clinicopathological and molecular data. Since our stage IV patient group consists of n= 511 patients for which we manually extracted data, we chose to focus our study on this (most interesting) stage IV group. Our research objective was to model the associations between the risk of having an SD versus FIT-IC stage IV CRC and the identified (significantly) different characteristics, called variables hereafter.

METHODS

Study population and study design

This retrospective population-based study is a follow-up study of Tran *et al.* [7]. In the current study, all eligible individuals (between 53 through -74 years old) who participated in the Flemish CRC screening program between October 2013 (start of the program) and December 2018 (last year with all required data complete) and who were diagnosed with either a stage IV SD or FIT-IC CRC in the same period were selected. All cases were screened using another FIT (OC Sensor, Eiken, Japan). We adopt the same definitions as those used in the previous study for SD and FIT-IC CRC, where SD CRC was defined as "a CRC that was diagnosed within 6 months after the first follow-up colonoscopy for a positive FIT and before the next recommended FIT invitation (24 months)" [7]. A FIT-IC CRC was defined as "a CRC diagnosed after a negative FIT and before the next recommended FIT invitation (24 months)". Staging was performed using the applicable TNM edition at the time of diagnosis (TNM 7th edition for incidence years 2013-2016, TNM 8th edition starting from incidence year 2017). Combined TNM stage was a compilation of pathological (pTNM) and clinical (cTNM) stage. pTNM prevails over cTNM, except when cTNM stage is IV [10]. All available clinical, clinicopathological and molecular data was extracted from the Belgian Cancer Registry (BCR).

Data sources and studied variables

Data from all patients (n=511) was obtained from several databases. First, the Flemish Centre for Cancer Detection provided data on participant's screening history (FIT result and follow-up colonoscopy), which originates from linkage of data from the InterMutualistic Agency. Then, population-based data from the Belgian

Cancer Registry (BCR) was subsequently used to link these data to personal information (gender, age), tumor characteristics (location, adenocarcinoma type…), clinicopathological and molecular features (lymphovascular invasion, genetic mutations…). All variables are described in Table 1. For patients with multiple lesions (<5% of the total population), the lesions were considered independent. Tumor location was classified as right sided colon (from cecum to hepatic flexure), transverse colon or left sided colon (from splenic flexure to the sigmoid). In our population, there were no patients with stage IV rectal tumors. Tumor types were split in adenocarcinomas ('no specific type (NST)', mucinous and signet ring cell carcinomas) and neuroendocrine tumors (NET). Tumor sizes were grouped per 2 cm. Tumor descriptions were summarized as bulky (including polypoid, bourgeois and stenosing tumors), ulcerative (comprising flat, ulcerating and infiltrating tumors) or both (combinations of the former). The presence or absence of another primary tumor was reported as such. In case a significant result was found, additional analysis for the location of the other primary tumor was performed. Lymphovascular invasion (LVI) was subdivided in being absent (none), only lymphatic invasion, only vascular invasion and both lymphatic/vascular invasion. Depth of Invasion (Di) was subdivided in T0-2, T3 and T4. Microsatellite instability (MSI) was categorized as stable, low or high. DNA mismatch repair genes were documented in view of the protein staining and were described as positive when a loss of staining was found.

Statistical analyses

Sample size and missing data

We included all 507 eligible stage IV CRCs diagnosed tumors in the Flemish CRC screening program between October 2013 and December 2018. For a few patients, there was no information about the tumour; these cases were removed. Data on gender, age and other primary tumors was complete (see Table 1). Data for location was complete for 97% and type of adenocarcinoma for 99% of all cases. Table 1 gives an overview of the number of cases for which data on the different characteristics were available.

Table 1 | Study population characteristics

Parameters	**Outcome – split IC/SD**		**Total n**
Outcome	**IC, n (%)**	**SD, n (%)**	
	215 (42.2%)	292 (57.6%)	507
Gender, female	112 (52.1%)	103 (35.3%)	507
Location			
Right	106 (51.2%)	66 (23.1%)	493
Left	85 (41.1%)	210 (73.4%)	
Transverse	16 (7.7%)	10 (3.5%)	
Age (years)			
Median (IQR)	66 (62-71)		507
50-54	0 (0.0%)	3 (1.0%)	
55-59	31 (14.4%)	43 (14.7%)	
60-64	45 (20.9%)	69 (23.6%)	
65-69	58 (27.0%)	82 (28.1%)	

70-74	60 (27.9%)	92 (31.5%)	
75-79	21 (9.8%)	3 (1.0%)	
Type of adenocarcinoma*			
NST	161 (75.6%)	247 (84.9%)	504
Mucinous	37 (17.4%)	33 (11.3%)	
Signet ring cell	15 (7.0%)	11 (3.8%)	
NET (presence)	12 (5.6%)	4 (1.4%)	505
Tumor size (cm)			
Median (IQR)	3.75 (2.35 – 5.75)		303
0 – 2.0	12 (12.5%)	33 (16.4%)	
2.1 – 4.0	51 (53.1%)	101 (50.2%)	
4.1 – 6.0	26 (27.1%)	50 (24.9%)	
6.1 – 8.0	5 (5.2%)	12 (6.0%)	
8.1 – 10.0	3 (3.1%)	3 (1.5%)	
10.1 – 12.0	3 (3.1%)	2 (1.0%)	
12.1 – 14.0	1 (1.0%)	0 (0.0%)	
16.1 – 18.0	1 (1.0%)	0 (0.0%)	
Tumour description			
Bulky	21 (22.1%)	23 (14.2%)	257
Ulcerative	61 (64.2%)	103 (63.6%)	
Both	13 (13.7%)	36 (22.2%)	
Lymph node metastasis (presence)	93 (77.5%)	151 (72.2%)	329
Lymphovascular invasion			
None	30 (25.6%)	85 (42.9%)	315
Only lymphatic invasion	10 (8.5%)	18 (9.1%)	
Only vascular invasion	15 (12.8%)	19 (9.6%)	
Both lymphatic and vascular invasion	62 (53.0%)	76 (38.4%)	
Depth of invasion			
T0-2	9 (7.8%)	23 (11.3%)	320
T3	55 (47.4%)	127 (62.3%)	
T4	52 (44.8%)	54 (26.5%)	
Perineural invasion (presence)	39 (39.8%)	62 (36.7%)	267
Extra tumoral deposits (presence)	31 (60.8%)	41 (55.4%)	125
MSI			
Stable	77 (95.1%)	101 (54.3%)	186
Low	2 (2.5%)	1 (0.9%)	
High	2 (2.5%)	3 (2.8%)	
MLH1 (positive)	105 (99.1%)	146 (97.3%)	259
PMS2 (positive)	104 (95.4%)	138 (97.2%)	251
MSH2 (positive)	106 (99.1%)	149 (100%)	256
MSH6 (positive)	106 (97.2%)	147 (100%)	256
MLH1 methylation (presence)	1 (100.0%)	1 (100.0%)	2
APC (positive)	11 (84.6%)	6 (85.7%)	20
KRAS (positive)	57 (47.1%)	89 (48.6%)	304
NRAS (positive)	8 (8.3%)	7 (5.2%)	230
HRAS (positive)	0 (0.0%)	0 (0.0%)	78
BRAF (positive)	16 (18.2%)	13 (13.1%)	187
PIK3CA (positive)	12 (27.9%)	4 (8.3%)	91
Other primary tumor (presence)	35 (16.8%)	35 (12.0%)	507

IC= Interval cancer, SD= screen-detected, IQR= inter quartile range, NST= no specific type, NET= neuro-endocrine tumor, MSI= microsatellite instability

* Adenosquamous and medullary adenocarcinoma were left out because there were too few cases.

Main analyses

All categorical variables were described as counts and percentages. Continuous variables were described with their median and interquartile ranges. To evaluate whether significant differences were present between SD or IC CRC per characteristic, we performed either a Welch's t-test for continuous variables or a chi-square/fisher exact test for categorical variables. P-values below 0.05 (two-sided) were considered to be statistically significant. For variables consisting of multiple levels (e.g. type of adenocarcinoma), a Benjamini-Hochberg correction for multiple testing was implemented. P-values less than the adjusted threshold based on a 0.1 false discovery rate (FDR) were considered to be statistically significant. A higher FDR of 0.1 was used because of the exploratory character of the analyses. The Benjamini-Hochberg correction is demonstrated in Supplementary table 1.

To identify variables associated with the odds of having a stage IV SD or FIT-IC CRC, logistic regression was performed. All variables that were significantly different in the exploratory analyses, were first tested in a univariate model. Crude odds ratios with 95% confidence intervals were reported. A likelihood ratio test was performed with null hypothesis that all categories carry the same odds to have a FIT-IC. Benjamini-Hochberg correction was applied to correct for multiple testing. P-values below the adjusted threshold based on a 0.05 false discovery rate (FDR) were considered to be statistically significant. In case the independent variable had more than two levels, post hoc analysis with Dunnett correction was performed.

The previous study by Tran *et al.* [7] found significant associations between age, gender, and location and the risk of having FIT-IC vs SD-CRC, therefore we included these variables as covariates in multivariable analyses in the current study. Multicollinearity between the covariates and other independent variables was checked and only reported if present. Adjusted odds ratios with 95% confidence intervals were reported. P-values below the adjusted threshold based on a 0.05 false discovery rate (FDR) were considered to be statistically significant. Benjamini-Hochberg correction is demonstrated in Supplementary table 2.

Privacy, ethical approval and consent to participate

When participating in the Flemish CRC-SP, all participants filled out a written informed consent explaining that personal information can be used for scientific research and evaluation to improve the CRC screening program. In this study, data from the Flemish Centre for Cancer Detection and Belgian Cancer Registry was used, for which approval was given by the Belgian Privacy Commission (reference IVC/KSZG/19/236, number 13/091 [11]). All data was pseudonymized. The study was conducted in accordance with the Declaration of Helsinki.

RESULTS

Study population

A total of 507 stage IV CRCs was included, 215 (42.2%) of which were diagnosed with a FIT-IC and 292 (57.6%) with an SD CRC. Most of the FIT-ICs were diagnosed among females (52.1%) with right sided tumors (51.2%) at older age (70-74; 27.9%), as reported before [7]. Overall, most of the tumors were adenocarcinomas with no specific type (NST, 81.0%). For 8 out of 27 variables, data was >90% complete. For 10 out of 27 variables, data was >50% complete. For 9 out of 27 variables, data was less than 50% complete. For 4 variables, data was less than 20% complete (see Table 1).

Exploratory analyses of clinicopathological and molecular features in SD vs FIT-IC CRC

Table 2 gives an overview of all statistical analyses and the corresponding P-values. For gender, location, NET and presence of a *PIK3CA* mutation a significant difference in SD vs FIT-IC was found. For gender and location, this is in line with the previous study [7]. NET and *PIK3CA* presence were more frequent for FIT-IC CRC. Adenocarcinoma type, depth of invasion and lymphovascular invasion also showed significant differences between SD and FIT-IC, and since these variables had multiple categories, further testing was performed. We found significant differences between T0-2 vs T4 and T3 vs T4 of SD vs FIT-IC CRC's depth of invasion, where T4 was more frequent in FIT-ICs but T0-2 and T3 more frequent in SD CRC. Also, there was a significant difference found for (lympho)vascular invasion compared to no invasion for SD vs FIT-IC CRC. Here, lymphovascular invasion was more common in FIT-IC, while no invasion was more frequent in SD CRC. For type of adenocarcinoma, NST vs mucinous was significant for SD vs FIT-IC with mucinous adenocarcinoma being more frequent in FIT-ICs and adenocarcinomas of NST more frequent in SD CRC. Age, tumor size and MSH6 mutations only show a trend towards significance (p-value < 0.1).

Table 2 | Exploratory analyses of SD vs FIT-IC colon cancers

Variable	P-value	P-value Post hoc[1]
Gender[2]	**0.000219**	/
Age[3]	0.0757	/
Location	**4.32E-12**	Right vs Left: **3.48 E-12** Left vs Transverse: **0.000572**
Adenoma type	**0.0291**	NST vs mucinous: **0.0355***
NET	**0.0152**	/
Tumor size	0.0790	/
Tumor description	0.106	/
LNM	0.359	/
Depth of invasion	**0.00352**	0-2 vs 4: **0.0367** T3 vs T4: **0.00142**
Perineural invasion	0.708	/
Lymphovascular invasion	**0.0159**	None vs only vascular: **0.0443**

		None vs both: **0.00192**
ETD	0.679	/
MSI	0.847	/
MLH1_meth	NA	/
MLH1	0.731	/
PMS2	0.494	/
MSH2	0.411	/
MSH6	0.0795	/
APC	1	/
KRAS	0.809	/
NRAS	0.435	/
HRAS	NA	/
BRAF	0.426	/
PIK3CA	**0.0270**	/
Another primary tumor	0.196	/

NST= no specific type, NET= neuro-endocrine tumor, LNM= lymph node metastasis, ETD= Extra Tumoral Deposits, MSI= microsatellite instability

[1] Multiple testing correction (Benjamini-Hochberg) was used. Only significant p-values are reported in this table in bold. More details of the adjusted p-values can be found in Supplementary Table 1.

[2] Gender was tested as one variable. Previous analyses already demonstrated that FIT-ICs are more common in women

[3] All ages were tested together. Previous analyses already demonstrated that older ages give increased odds for FIT-IC

Clinicopathological and molecular features associated with the risk of having a FIT-IC vs SD CRC

Univariable logistic regression

For each variable that was found to be significant in the exploratory analyses, the odds of having a FIT-IC vs SD CRC was modeled using a univariable logistic regression. Table 3 gives an overview of all tested variables, with p-values and crude odds ratios. Within the univariate models, NET presence, lymphovascular invasion and *PIK3CA* mutation were significant. FIT-IC CRCs have a 4-fold increased odds to be a NET tumor (OR= 4.26 [1.46 – 15.40]), a 2-fold increased odds to present with lymphovascular invasion (OR= 2.31 [1.36 – 3.98]) and a 4-fold increased odds to have a *PIK3CA* mutation compared to SDCRC (OR= 4.25 [1.34 – 16.40]). The mucinous subtype of adenocarcinoma and depth of invasion T4 gave a close to significant result compared to NST and T0-2 respectively ($p < 0.1$).

Table 3 | Logistic regression for variables with univariate associations. Univariate models are displayed, as well as multivariate models corrected for at least age, gender and location.

Characteristic	Category	Crude OR (95% CI)	P-value	aOR (95% CI)	P-value[1]
Type of adenocarcinoma	NST	ref	0.0303	ref	0.747
	Mucinous	1.72 [1.03 – 2.87]	0.0724	1.16 [0.66 – 2.04]	/
	Signet ring cell	2.09 [0.94 – 4.78]	0.1377	1.28 [0.53 – 3.15]	/
NET	Absence	ref	**0.00709**	ref	**0.00336**
	Presence	4.26 [1.46 – 15.40]	NA	5.29 [1.71 – 20.00]	NA
Depth of Invasion	T0-2	Ref	**0.00382**	Ref	0.137

	T3	1.10 [0.49 – 2.67]	0.939	1.00 [0.41 – 2.62]	/
	T4	2.46 [1.07 – 6.07]	0.0616	1.66 [0.63 – 4.56]	/
Lymphovascular invasion	No	Ref	**0.0144**	Ref	**0.0244**
	Only lymphatic	1.57 [0.64 – 3.74]	0.645	1.42 [0.54 – 3.55]	0.949
	Only vascular	2.23 [1.00 – 4.96]	0.127	1.49 [0.60 – 3.61]	0.696
	Both	2.31 [1.36 – 3.98]	**0.00623**	2.50 [1.41 – 4.52]	**0.0151**
PIK3CA	Absence	Ref	**0.0130**	Ref	0.163
	Presence	4.25 [1.34 – 16.40]	NA	2.6 [0.68 – 11.20]	NA

[1] P-values next to the reference level refer to a likelihood ratio test, with null hypothesis that all categories carry the same odds to have a FIT-IC. P-values in bold survive the multiple hypothesis correction (Benjamimi-Hochberg, suppl Table 2). In case the independent variable has more than two levels, the P-values next to the non-reference levels refer to the Dunnett-corrected P-values from the post hoc analysis, comparing the non-reference level to the reference.

Multivariable logistic regression

Since the previous work showed associations between gender, age and location and the risk of having a FIT-IC vs SD CRC, we included these variables as covariates in the multiple logistic regression models [7]. Table 3 gives an overview of all tested variables, with p-values and adjusted odds ratios. Figure 1 gives an overview of the adjusted odds ratios for the tested variables. In the multivariate models, only NET presence and lymphovascular invasion showed significant results. In the corrected models, FIT-IC CRC have 5 times higher odds to be a NET tumor (OR= 5.29 [1.71 – 20.00]) and 2.5 times higher odds to have lymphovascular invasion (OR= 2.50 [1.41 – 4.52]). All other variables completely lost significance when the models were corrected for age, gender and location.

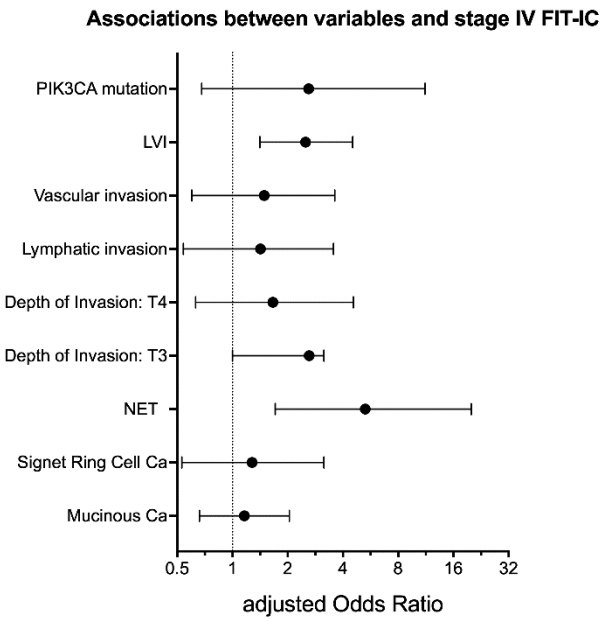


Figure 1 | Adjusted Odds Ratio with 95% confidence intervals for variables associated with stage IV FIT-IC. LVI= lymphovascular invasion, NET = neuro-endocrine tumor, Ca= carcinoma. Figure created with Graphpad Prism v10.1.1

DISCUSSION

In this study, we identified differences between clinicopathological and molecular characteristics of SD and FIT-IC and assessed their associations with the risk of having a FIT-IC vs SD CRC.

In the exploratory analyses, we identified significant differences between SD versus FIT-IC for gender, presence of NET, depth of invasion (T3 vs T4), lymphovascular invasion (none vs both) and presence of a *PIK3CA* mutation. For 'type of adenocarcinoma', significant differences were found, but significance was lost when testing the different types separately, partly due to the Benjamini Hochberg correction. Regression analyses of FIT-IC versus SD stage IV CRC revealed new interesting insights about the significant parameters and their association with higher risk of having a FIT-IC. All relevant factors are discussed below.

Previously found variables: gender, age and location

Consistent with the previous study by Tran *et al.* [7], there is a significant difference in prevalence of SD versus FIT-IC CRC between genders, with FIT-ICs more likely to occur in women. Reasons reported for this are a lower blood hemoglobin concentration, a longer colonic transit time and a higher proportion of right-sided tumors [7]. It has been suggested that women are more likely to seek medical help when they experience symptoms, which may lead them to undergo diagnostic tests and, as a result, have their CRC diagnosed as FIT-IC [12,13]. Furthermore, we replicated the very significant effect of location, especially right vs left. An in-depth discussion on these differences was given previously [7]. In short, a longer transit time from the right side and higher proportion of flat tumors are thought to play a role herein. For these reasons, the current study included gender and location as covariates in the multiple regression analysis. Contrary to the previous study, we did not find a significant association between age nor age categories and the odds of having a FIT-IC versus a SD stage IV CRC. However, in this study we have only included stage IV SD and IC CRCs, which could explain these differences[7]. Nevertheless, age was also taken into account as a covariate in later analyses (multiple logistic regression).

Tumor Type

Within our study, we made the distinction between adenocarcinoma and NET colon tumors because of their different origin (epithelial vs neuroendocrine cells) and the rarity of NETs [14]. Adenocarcinoma tumors were split by per histological type, as described by the WHO [15].

In literature, few studies have investigated adenocarcinoma type in SD vs FIT-ICs. Steel *et al.* [16] did report an association of aggressive histotypes (mucinous and signet ring cell adenocarcinoma) with higher risk of FIT-IC. However, in our exploratory analyses, we only found significant results for mucinous adenocarcinomas. Signet

ring cell carcinoma lost its significance after Benjamini Hochberg correction. Within the logistic regression analysis, mucinous adenocarcinoma was only found close to significant after correction for multiple testing. Importantly, the multiple logistic regression analysis revealed that, after adjusting the model for location, the type of adenocarcinoma was not significant anymore, indicating that location, especially right-sided tumor location, is an important variable that needs to be taken into account in the analysis of interval CRCs. It has been reported that mucinous adenocarcinomas occur in 10-20% of CRC patients, mostly women, in the proximal colon (right and transverse) [17]. This is concordant with our population, where 13.6% cases were mucinous adenocarcinomas (both SD and FIT-IC), 57% of mucinous tumors being observed in female patients and 60% located in the right colon. Signet ring cell carcinoma was also reported to be more common in the proximal colon [18], which we also observed in our cohort (88.4% for both SD and FIT-IC). In all, our results show that tumor location is an important confounder, that needs to be accounted for in the analysis of the association between adenocarcinoma type and the odds of having a FIT-IC versus a SD CRC.

In both the exploratory and regression analyses, the presence of a NET tumor was found to be significantly different in FIT-IC vs SD CRC, with 5.3 times higher odds of presence of NET in FIT-ICs. A few reasons could explain this observation. First, NET tumors are completely different compared to adenocarcinoma tumors in origin and clinical presentation [19]. NETs tend to bleed less than adenocarcinomas [20], partly explaining the higher risk for FIT-IC, although 4 NETs were screen-detected in our cohort. However, we cannot be certain that for example, other lesions caused the bleeding and that the NET was accidently discovered during the colonoscopy This hypothesis was already described by others [21–23]. However, if this had been the case, the number of SD NET would have been smaller and the number of interval NET would have been larger than the numbers shown in this study, leading to a larger OR which indicates an even stronger association between NET and the odds of being a FIT IC versus a SD CRC. In our study, we found 1.4% of all SD cancers to be NETs. In literature, there are very few research papers describing NETs found after colonoscopy. Most of those papers describe its prevalence, which is ranging from 0.018% in the English bowel screening program [21], to 0.16% in the Taiwanese programme [24–27]. However, these studies almost exclusively focus on rectal NETs, which we do not have in our cohort. Moreover, other studies take into account all stages. Therefore, it is difficult to compare our results to the existing literature. There is one study by Kim et al. that focuses on colon subepithelial tumors discovered by chance. Here, of the 105 detected tumors, 2 were NETs (1.9%), which is close to our rate of 1.4% [28]. Overall, it is important to keep in mind that colorectal NET tumors are difficult to find with FIT screening, and other options will need to be explored for NET detection in the colon.

Clinicopathological variables

Several clinicopathological variables were investigated in our study. They can be grouped by tumor-informed parameters (lymphovascular invasion (LVI), perineural invasion (PNI), depth of invasion (Di), lymph node

metastasis (LNM) and extra tumoral deposits (ETD)) and tumor-agnostic parameters (microsatellite instability (MSI) and DNA mismatch repair (MMR) markers for MSI).

Regarding the tumor-informed parameters, only Di and LVI were found to be significant in our exploratory analysis. When further investigating these variables, we found associations between LVI and PNI, Di and LNM. Remarkably, Di was not associated with PNI. ETDs were not associated with LVI or Di. The associated variables were taken into account when performing the multiple regression analyses. In the final multivariable logistic regression analysis, only presence of LVI was found to be a significantly associated risk factor for FIT-IC vs SD, with almost 2 times higher odds of being a FIT IC versus a SD CRC. In literature, there is no information that describes the role of any of the clinicopathological variables in CRC FIT-ICs. The only relevant information is that LVI and PNI have been described as independent negative prognostic factors (i.e. poor outcome) in CRC development [29–31]. There is one group that also describes LVI as a risk factor for developing metastasis in CRC [32], which underlines the importance of reporting LVI at diagnosis. Interestingly, presence of LVI was also described as a significantly factor in a study about SD vs IC in breast cancers, with LVI more present in IC [33]. Together with our results, this shows an important role for LVI in several malignancies, but further research is needed to fully understand its role in (FIT-) ICs.

Regarding the tumor-agnostic clinicopathological features, we did not find any significant difference in our analyses. Only in the exploratory analysis, a positive MSH6 result was found to be close to significantly different in SD vs FIT-IC cancers (p= 0.08). In literature, there are only a few papers describing MSI in ICs, and the findings are contradictory. Agreeing with our findings, Soong *et al.* did not find any difference in MMR expression between IC and SD CRC [34]. However, contrary to our research, their population was screened using colonoscopies and not FITs. The ICs were defined as "cancer detected in a diagnostic examination prior to the next recommended colonoscopy and at least 1 year after the last colonoscopy". The gap between these screening tools might explain the differences found in the variables. A few other papers describe that, despite the insignificant results, MSI is observed to be more prevalent in ICs [35–37]. We did not observe this (Table 1). Lastly, two research papers describe significant differences in MSI for SD vs IC, but they defined IC patients as "Subjects with 1 prior colonoscopy > 180 days before the diagnosis" and "individuals that had a complete colonoscopy performed within 5 years of the diagnosis of CRC" respectively, which is different from the definition in more recent literature [38,39]. Also, the sample size of the latter was small (n= 42) [39]. There is clearly still no consensus about the role of MSI in (FIT-)ICs and despite the predictive role of MSI in CRC outcome [40], we cannot conclude if there is a role of MSI in the development of FIT-ICs.

Genetic alterations

The most commonly reported genes were investigated in our study. We found significant differences between SD and FIT-IC CRC for the presence of a *PIK3CA* mutation in the exploratory and close to significant results in the simple logistic regression analysis, with an OR of 4.25 in the latter.

When performing the multiple logistic regression analysis, the significance for *PIK3CA* was lost. In literature, there are no papers describing *PIK3CA* mutations in FIT-ICs. There are few papers describing *PIK3CA* mutations in a colonoscopy-screened population, where in concordance to our multivariate regression results, no significant differences were found between SD and IC CRC. However, sample sizes might have been too small in both our (for *PIK3CA*: n=92) and these other studies to detect less frequent genetic differences [34,39,41]. Furthermore, several studies describe that *PIK3CA* mutations are more often present in right-sided colon tumors [42–45], which could explain why this variable lost significance after adjusting for tumor location. Nevertheless, the use of *PIK3CA* mutations as biomarkers remains a discussion point in literature. It has been described as a prognostic biomarker for aggressive tumor growth and increased risk of tumor recurrence [45]. Furthermore, associations between FIT-IC and other genetic alterations in e.g. *KRAS* have been reported [45], although we did not find this in our analyses. All other genes that were studied in our cohort, did not show any significant difference between SD and FIT-IC CRCs. This is in line with what is described in literature [34,39,41,46], although these studies do not involve FIT-screened patients but on colonoscopy-, flexible sigmoidoscopy- or gFOBT- screened patients.

Missing data, difficulties and limitations of this study

One of the difficulties of working with the pathology reports from different hospitals and different years (2013 to 2018) is the inconsistent reporting of different features. For some variables, such as tumor type, location, age and gender, the reports were almost complete (>97%), while for example the tumor-informed clinicopathological parameters were only complete for ~50% of all cases. This means different (sub)datasets had to be used for the analysis of each variable. In this view, we also performed the analyses with the subset of data where all variables were complete (total n= 247). Despite the loss of statistical significance in this analysis, similar trends were observed for NET, LVI and *PIK3CA*. We cannot distinguish if missing data is coming from inconsistent reporting or inconsistent testing. Although guidelines for testing e.g. MSI and molecular alterations do exist, implementation is not perfect. Moreover, there is no standardization of the pathology reports between Belgian hospitals. Lastly, additional testing, especially for molecular markers, is often reported in additional reports rather than added to the original report, where we extracted the data from. All these hurdles made it difficult to obtain 100% complete datasets in this study. However, missing is at random, therefore it is highly unlikely that this missingness would affect our conclusions.

For some parameters, e.g. *APC* and *MLH1* methylation, there was a lack of sufficient data for statistical analysis, as such the relevance of these alterations in FIT-IC could not be assessed. Overall, it was difficult to perform reliable analyses for molecular markers, mainly because of a small sample size for molecular alterations (only 18 - 60% data completeness). By the end of 2014, the European Society of Medical Oncology (ESMO) published evidence-based guidelines for molecular testing of specific genes of the *EGFR* pathway in metastatic CRC [47]. This pathway contains important actionable targets for selection of first-line therapy [48,49]. Furthermore, *KRAS/NRAS* and *BRAF* are considered important predictive and prognostic biomarkers for treatment decisions in metastatic CRC [48]. Our cohort data is partly coming from a period before this year, which can – to a certain extent - explain the lack of completeness in reporting of the mutations. However, this underlines the need to work towards more complete information. Up until today, there is still not enough evidence to use *PIK3CA* outside of clinical trials [49], which could explain why mutations in this gene are not frequently tested or reported. As a final remark on molecular alterations, the lack of significant findings in molecular alterations could also be because only stage IV patients were included. This lowered the number of cases and thus the chances to find significant associations with the study outcome, if there are any, for (less frequent) mutations.

Future perspectives

Despite the advantages of CRC screening there are still some important limitations in using FIT [50]. For example, in Flanders, participation rates are suboptimal compared to levels recommended by the European commission [51], fluctuating between 48.0% and 52.5%, with a lower participation rate in the younger age groups (50-54 years). Around 1 in 5 non-responders have reported their reason for non-participation being either fear of a FIT false positive result or a dislike for the procedure of fecal testing [52]. Therefore, other screening methods, including blood-based biomarkers, have gained more attention in recent years [53].

Very recently, the HUNT study of Brenne *et al*. showed that CRC can be detected up to 2 years prior to clinical diagnosis, based on methylated circulating tumor DNA (ctDNA) [54]. This research suggests that patients could receive their diagnosis up to 2 years earlier than the clinical diagnosis if ctDNA analysis would be part of the CRC screening program. This also leads us to believe that FIT-IC proportions - particularly in stage IV - could potentially be lowered using ctDNA analysis. However, studies show that existing ctDNA-based tests, e.g. EpiProColon® and Galleri®, are not cost-effective at their current cost ($192 and $950 respectively) and screening performance. Liquid biopsy testing for CRC could potentially become more cost-effective than FIT, but only if the cost is substantially lowered. Further clinical trials are needed to investigate also the uptake in large-scale population screening, as the current ctDNA tests are not yet suited for this purpose [45,46].

In view of our results, liquid biopsy-based screening could for example indicate the presence of a NET tumor, based on NET-specific markers [55]. Although rare, colorectal NETs are a tumor type that is often missed by the

FIT and mostly detected by accident when performing a colonoscopy. In the future, larger analyses for molecular alterations could prove useful. Moreover, a trend towards epigenetic research could be followed by also investigating DNA methylation biomarkers for FIT-IC. In the future, more comprehensive reporting – albeit more consistent reporting or more consistent testing- should also be considered. Lastly, this research could be expanded towards all stages to find critical characteristics that could lead to diagnosis of earlier stage CRC.

CONCLUSION

In this study, we evaluated clinicopathological and molecular difference between SD vs FIT-IC stage IV CRC. Throughout all analyses, the presence of NET and lymphovascular invasion were newly identified as factors associated with higher odds of having a stage IV FIT-IC instead of a stage IV SD CRC. Further research will be needed to clarify how these insights might help in optimizing the Flemish CRC screening program. Besides these observations, we found that tumor location is a crucial covariate when analyzing clinicopathological and molecular factors in FIT-ICs. Therefore, tumor location should be taken into account (where applicable) in the analyses concerning FIT-ICs. Lastly, expanding the study to all stages and prospective validation of these and future results will be necessary before potentially implementing it into the program and as such, optimizing CRC screening.

DECLARATIONS

Author contributions

IN: Conceptualization of the study, methodology, interpreting results, writing original draft, reviewing and editing of the final draft.

NTN: Conceptualization of the study, methodology, interpreting results, reviewing and editing of the final draft.

AF: Conceptualization of the study, interpreting results, reviewing and editing of the final draft.

SJ: Interpreting results, reviewing and editing of the final draft.

KVH: Interpreting results, reviewing and editing of the final draft.

EF: Methodology, interpreting results, reviewing and editing the final draft.

MP: Reviewing and editing of the final draft.

KODB: Interpreting results, reviewing and editing the final draft.

GVC: Interpreting results, reviewing and editing of the final draft.

SH: Conceptualization of the study, interpreting results, reviewing and editing the final draft.

GVH: Conceptualization of the study, interpreting results, reviewing and editing of the final draft.

All authors have read and agreed to the published version of the manuscript.

Funding

Research in the lab of Prof. Guy Van Camp is supported by grants awarded by the University of Antwerp (IOF/SBO 43782, BOF/Methusalem grant 40790, BOF/TOP 39705). Isabelle Neefs is supported by a strategic basic PhD fellowship of the Research Foundation Flanders (FWO; IS67523N).

Competing interests

The authors declare no competing interests

Ethics approval and consent to participate

When participating in the Flemish CRC-SP, all participants filled out a written informed consent explaining that personal information can be used for scientific research and evaluation to improve the CRC screening program. In this study, data from the Flemish Centre for Cancer Detection and Belgian Cancer Registry was used, for which approval was given by the Belgian Privacy Commission (reference IVC/KSZG/19/236, number 13/091). All data was pseudonymized. The study was conducted in accordance with the Declaration of Helsinki.

REFERENCES

1. Belgian Cancer Registry. Cijfers over kanker. https://kankerregister.org/Cijfers_over_kanker. 2021.

2. Belgian Cancer Registry. Cancer Fact Sheet Colorectal Cancer 2021 [Internet]. Available from: https://kankerregister.org/media/docs/CancerFactSheets/2021/Cancer_Fact_Sheet_ColorectalCancer_2021.pdf

3. Balchen V, Simon K. Colorectal cancer development and advances in screening. Clin Interv Aging. 2016;Volume 11:967–76.

4. CENTRUM VOOR KANKEROPSPORING SK. Het Bevolkingsonderzoek Dikke Darmkanker in Vlaanderen.

5. Joint Statement Call for action on Colorectal Cancer Screening in the EU.

6. Hoeck S, Pringels S, Kellen E, Van Herck K, Martens P, Van Limbergen E, et al. First results of the Flemish colorectal cancer screening program : start-up- period late 2013. Acta Gastroenterol Belg. 2016;79:421–8.

7. Tran TN, Peeters M, Hoeck S, Van Hal G, Janssens S, De Schutter H. Optimizing the colorectal cancer screening programme using faecal immunochemical test (FIT) in Flanders, Belgium from the "interval cancer" perspective. Br J Cancer. 2022;126:1091–9.

8. van de Veerdonk W, Hoeck S, Peeters M, Van Hal G, Francart J, De Brabander I. Occurrence and characteristics of faecal immunochemical screen-detected cancers vs non–screen-detected cancers: Results from a Flemish colorectal cancer screening programme. United European Gastroenterol J [Internet]. 2020 [cited 2023 Dec 18];8:185. Available from: /pmc/articles/PMC7079275/

9. Niedermaier T, Balavarca Y, Brenner H. Stage-Specific Sensitivity of Fecal Immunochemical Tests for Detecting Colorectal Cancer: Systematic Review and Meta-Analysis. American Journal of Gastroenterology. Wolters Kluwer Health; 2020. p. 56–69.

10. James D. Brierley MKGCW. TNM Classification of Malignant Tumours, 8th Edition. Wiley-Blackwell; 2016.

11. Informatieveiligheidscomité - Kamer sociale zekerheid en gezondheid [Internet]. [cited 2023 Oct 30]. Available from: https://www.ehealth.fgov.be/ehealthplatform/file/view/AWvhE2PqnF_Mkwg-mMCV?filename=13-091-n236-bevolkingsonderzoek%20dikkedarmkanker-gewijzigd%20op%202%20juli%202019.pdf

12. Ballering A V, Hartman TCO, Verheij R, Rosmalen JGM. Sex and gender differences in primary care help-seeking for common somatic symptoms: a longitudinal study. [cited 2024 Aug 8]; Available from: https://doi.org/10.1080/02813432.2023.2191653.

13. Evans REC, Brotherstone H, Miles A, Wardle J. Gender differences in early detection of cancer. [cited 2024 Aug 8]; Available from: www.liebertpub.com

14. Fleming M, Ravula S, Tatishchev SF, Wang HL. Colorectal carcinoma: Pathologic aspects. J Gastrointest Oncol. Pioneer Bioscience Publishing; 2012. p. 153–73.

15. WHO Classification of Tumours Editorial Board. Digestive System Tumours: WHO Classification of Tumours. 5th ed. 2019.

16. Steel MJ, Bukhari H, Gentile L, Telford J, Schaeffer DF. Colorectal adenocarcinomas diagnosed following a negative faecal immunochemical test show high-risk pathological features in a colon screening programme. Histopathology. 2021;78:710–6.

17. Luo C, Cen S, Ding G, Wu W. Mucinous colorectal adenocarcinoma: Clinical pathology and treatment options. Cancer Commun. BioMed Central Ltd.; 2019.

18. An Y, Zhou J, Lin G, Wu H, Cong L, Li Y, et al. Clinicopathological and Molecular Characteristics of Colorectal Signet Ring Cell Carcinoma: A Review. Pathology and Oncology Research. Frontiers Media S.A.; 2021.

19. Osagiede O, Habermann E, Day C, Gabriel E, Merchea A, Lemini R, et al. Factors associated with worse outcomes for colorectal neuroendocrine tumors in radical versus local resections. J Gastrointest Oncol. 2020;11:836–46.

20. Cancer center. colorectal cancer types. https://www.cancercenter.com/cancer-types/colorectal-cancer/types. 2022.

21. Basuroy R, O'Donnell CM, Srirajaskanthan R, Ramage JK. Ileocolonic neuroendocrine tumours identified in the English bowel cancer screening programme. Colorectal Disease. 2018;20:O85–91.

22. Kooyker AI, Verbeek WHM, van den Berg JG, Tesselaar MET, van Leerdam ME. Change in incidence, characteristics and management of colorectal neuroendocrine tumours in the Netherlands in the last decade. United European Gastroenterol J. 2020;8:59–67.

23. Vootla V, Ahmed R, Niazi M, Balar B, Nayudu S. Synchronous Adenocarcinoma of the Colon and Rectal Carcinoid. Case Rep Gastroenterol. 2016;10:600–4.

24. Kooyker AI, Verbeek WH, van den Berg JG, Tesselaar ME, van Leerdam ME. Change in incidence, characteristics and management of colorectal neuroendocrine tumours in the Netherlands in the last decade. United European Gastroenterol J. 2020;8:59–67.

25. Lin C, Chung C, Chiang T, Tu C, Liang C. Detection of rectal neuroendocrine tumor during screening colonoscopy and its difference from colonic adenocarcinoma. Advances in Digestive Medicine. 2017;4:99–104.

26. Gallo C, Rossi RE, Cavalcoli F, Barbaro F, Boškoski I, Invernizzi P, et al. Rectal neuroendocrine tumors: Current advances in management, treatment, and surveillance. World J Gastroenterol. 2022;28:1123–38.

27. Nagtegaal ID, Vink-Börger E, Kuijpers CCHJ, Dekker E, Shepherd NA. Incidental findings in the bowel cancer population screening program: other polyps and malignancies – A nationwide study. Histopathology. 2023;82:254–63.

28. Kim A, Hong SN, Chang DK, Kim Y-H, Kim JE, Kim ER. Clinicopathologic and Endosonographic Characteristics of Colon Subepithelial Tumors Discovered Incidentally. Diagnostics. 2024;14:551.

29. Bianchi G, Annicchiarico A, Morini A, Pagliai L, Crafa P, Leonardi F, et al. Three distinct outcomes in patients with colorectal adenocarcinoma and lymphovascular invasion: the good, the bad, and the ugly. Int J Colorectal Dis. 2021;36:2671–81.

30. Al-Sukhni E, Attwood K, Gabriel EM, LeVea CM, Kanehira K, Nurkin SJ. Lymphovascular and perineural invasion are associated with poor prognostic features and outcomes in colorectal cancer: A retrospective cohort study. International Journal of Surgery. 2017;37:42–9.

31. Yahyazadeh H, Mafi AR, Khatooni E, Beheshti M, Abdollahinejad A. Lymphovascular and Perineural Invasions Are Independently Associated with Advanced Colorectal Carcinoma. Int J Cancer Manag. 2019;12.

32. Rönnow C-F, Arthursson V, Toth E, Krarup P-M, Syk I, Thorlacius H. Lymphovascular Infiltration, Not Depth of Invasion, is the Critical Risk Factor of Metastases in Early Colorectal Cancer. Ann Surg. 2022;275:e148–54.

33. Pálka I, Kelemen G, Ormándi K, Lázár G, Nyári T, Thurzó L, et al. Tumor Characteristics in Screen-Detected and Symptomatic Breast Cancers. Pathology & Oncology Research. 2008;14:161–7.

34. Soong TR, Nayor J, Stachler MD, Perencevich M, Jajoo K, Saltzman JR, et al. Clinicopathologic and genetic characteristics of interval colorectal carcinomas favor origin from missed or incompletely excised precursors. Modern Pathology. 2019;32:666–74.

35. van der Vlugt M, Carvalho B, Fliers J, Montazeri N, Rausch C, Grobbee EJ, et al. Missed colorectal cancers in a fecal immunochemical test-based screening program: Molecular profiling of interval carcinomas. World J Gastrointest Oncol. 2022;14:2195–207.

36. Lee YM, Huh KC. Clinical and biological features of interval colorectal cancer. Clin Endosc. Korean Society of Gastrointestinal Endoscopy; 2017. p. 254–60.

37. Dong SH, Huang JQ, Chen JS. Interval colorectal cancer: A challenging field in colorectal cancer. Future Oncology. Future Medicine Ltd.; 2018. p. 1307–16.

38. Stoffel EM, Erichsen R, Frøslev T, Pedersen L, Vyberg M, Koeppe E, et al. Clinical and Molecular Characteristics of Post-Colonoscopy Colorectal Cancer: A Population-based Study. Gastroenterology. 2016;151:870-878.e3.

39. Richter JM, Pino MS, Austin TR, Campbell E, Szymonifka J, Russo AL, et al. Genetic Mechanisms in Interval Colon Cancers. Dig Dis Sci. 2014;59:2255–63.

40. Gryfe R, Kim H, Hsieh ETK, Aronson MD, Holowaty EJ, Bull SB, et al. Tumor Microsatellite Instability and Clinical Outcome in Young Patients with Colorectal Cancer. New England Journal of Medicine. 2000;342:69–77.

41. Yang K, Cao Y, Gurjao C, Liu Y, Guo CG, Lo CH, et al. Clinical and Genomic Characterization of Interval Colorectal Cancer in 3 Prospective Cohorts. Gastroenterology. 2022;163:1522-1530.e5.

42. Ahn AR, Kim KM, Jang KY, Moon WS, Ha GW, Lee MR, et al. Correlation of PIK3CA mutation with programmed death ligand-1 (PD-L1) expression and their clinicopathological significance in colorectal cancer. Ann Transl Med. 2021;9:1406–1406.
43. Yamauchi M, Morikawa T, Kuchiba A, Imamura Y, Qian ZR, Nishihara R, et al. Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. Gut. 2012;61:847–54.
44. Rosty C, Young JP, Walsh MD, Clendenning M, Sanderson K, Walters RJ, et al. PIK3CA Activating Mutation in Colorectal Carcinoma: Associations with Molecular Features and Survival. PLoS One. 2013;8.
45. Vacante M, Borzì AM, Basile F, Biondi A. Biomarkers in colorectal cancer: Current clinical utility and future perspectives. World J Clin Cases. Baishideng Publishing Group Co; 2018. p. 869–81.
46. Ibáñez-Sanz G, Sanz-Pamplona R, Garcia M, behalf on, Beaulieu J-F, Binefa G, et al. cancers Future Prospects of Colorectal Cancer Screening: Characterizing Interval Cancers. Cancers (Basel). 2021;13:1328.
47. Van Cutsem E, Cervantes A, Nordlinger B, Arnold D, The ESMO Guidelines Working Group. Metastatic colorectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2014;25:iii1–9.
48. Sepulveda AR, Hamilton SR, Allegra CJ, Grody W, Cushman-Vokoun AM, Funkhouser WK, et al. Molecular Biomarkers for the Evaluation of Colorectal Cancer. The Journal of Molecular Diagnostics. 2017;19:187–225.
49. Cervantes A, Adam R, Roselló S, Arnold D, Normanno N, Taïeb J, et al. Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up ☆. Annals of Oncology. 2023;34:10–32.
50. Niederreiter M, Niederreiter L, Schmiderer A, Tilg H, Djanani A. Colorectal cancer screening and prevention—pros and cons. Memo - Magazine of European Medical Oncology. Springer-Verlag Wien; 2019. p. 239–43.
51. Segnan N (Nereo), Patnick Julietta, Karsa L von (Lawrence von), European Commission. Directorate General for Health & Consumers., International Agency for Research on Cancer. European guidelines for quality assurance in colorectal cancer screening and diagnosis. Publications Office of the European Union; 2010.
52. Hoeck S, Tran TN. Self-Reported Reasons for Inconsistent Participation in Colorectal Cancer Screening Using FIT in Flanders, Belgium. Gastrointestinal Disorders. 2023;5:1–14.
53. Ferrari A, Neefs I, Hoeck S, Peeters M, Van Hal G. Towards novel non-invasive colorectal cancer screening methods: A comprehensive review. Cancers (Basel) [Internet]. 2021 [cited 2023 Nov 6];13:1820. Available from: https://www.mdpi.com/2072-6694/13/8/1820/htm
54. Brenne SS, Madsen PH, Pedersen IS, Hveem K, Skorpen F, Krarup HB, et al. Colorectal cancer detected by liquid biopsy 2 years prior to clinical diagnosis in the HUNT study. Br J Cancer. 2023;129:861–8.
55. Mariën L, Islam O, Chhajlani S, Lybaert W, Peeters M, Van Camp G, et al. The Quest for Circulating Biomarkers in Neuroendocrine Neoplasms: a Clinical Perspective. Curr Treat Options Oncol [Internet]. 2023; Available from: https://link.springer.com/10.1007/s11864-023-01147-3

SUPPLEMENTARY MATERIAL

Supplementary Table 1 | Adjusted P-values from the Benjamini – Hochberg Post-hoc analysis of the exploratory analysis

Characteristic	P value	Rank (i)	Tests performed (m)	FDR (Q)	(i/m)*Q
Right vs Left	3.47E-12	1	15	0.1	0.0067
Left vs Transverse	0.000572	2	15	0.1	0.0133
T3 vs 4T	0.00142	3	15	0.1	0.0200
No LVI vs LVI	0.00192	4	15	0.1	0.0267
NST vs Mucinous Ca	0.0355	5	15	0.1	0.0333
T0-2 vs T4	0.0366	6	15	0.1	0.0400
No LVI vs only vascular	0.0443	7	15	0.1	0.0467
NST Ca vs signet ring cell Ca	0.0664	8	15	0.1	0.0533
No LVI vs only lymphatic	0.309	9	15	0.1	0.0600
Only lymphatic vs LVI	0.370	10	15	0.1	0.0667
Only llymphatic vs only vascular invasion	0.502	11	15	0.1	0.0733
Mucinous Ca vs signet ring cell Ca	0.677	12	15	0.1	0.0800
T0-2 vs T3	0.811	13	15	0.1	0.0867
Vascular vs LVI	0.932	14	15	0.1	0.0933
Right vs Transverse	0.993	15	15	0.1	0.1000

M= number of comparisons performed (15). Q= False Discovery Rate (0.1). LVI= lymphovascular invasion. NST= no special type. Ca= Carcinoma

Supplementary Table 2 | Adjusted P-values from the Benjamini – Hochberg correction of the logistic regression analyses

Analysis	Characteristic	P value	tests performed (m)	Rank (i)	FDR (Q)	(i/m)*Q
Multivariate	NET	0.00336	10	1	0.05	0.0050
Univariate	Depth of Invasion	0.00382	10	2	0.05	0.01000
Univariate	NET	0.00709	10	3	0.05	0.01500
Univariate	PIK3CA mutation	0.0130	10	4	0.05	0.0200
Univariate	Lymphovascular invasion	0.0144	10	5	0.05	0.0250
Multivariate	Lymphovascular invasion	0.0244	10	6	0.05	0.0300
Univariate	Type	0.0303	10	7	0.05	0.0350
Multivariate	Depth of Invasion	0.137	10	8	0.05	0.0400
Multivariate	PIK3CA mutation	0.163	10	9	0.05	0.0450
Multivariate	Type	0.747	10	10	0.05	0.0500

M= number of comparisons performed (30). Q= False Discovery Rate (0.05). LVI= lymphovascular invasion. NST= no special type. Ca= Carcinoma

CHAPTER 4

Epigenome-wide methylation analysis of colorectal carcinoma, adenoma and normal tissue reveals novel biomarkers addressing unmet clinical needs

(published in *Clinical Epigenetics* – DOI: 10.1186/s13148-023-01516-7)

Katleen Janssens[1,2*], Isabelle Neefs[1,2*], Joe Ibrahim[1,2*], Anne Schepers[1], Patrick Pauwels[2], Marc Peeters[2], Guy Van Camp[1,2*] and Ken Op de Beeck[1,2*]

**Katleen Janssens, Isabelle Neefs and Joe Ibrahim have contributed equally to this work (shared first authors), as well as Guy Van Camp and Ken Op de Beeck (shared last authors)*

1 Centre of Medical Genetics, University of Antwerp and Antwerp University Hospital, Prins Boudewijnlaan 43, 2650 Edegem, Belgium

2 Centre for Oncological Research Antwerp (CORE), University of Antwerp and Antwerp University Hospital, Universiteitsplein 1, 2610 Wilrijk, Belgium

ABSTRACT

Background: Biomarker discovery in colorectal cancer has mostly focused on methylation patterns in normal and colorectal tumor tissue, but adenomas remain understudied. Therefore, we performed the first epigenome-wide study to profile methylation of all three tissue types combined and to identify discriminatory biomarkers.

Results: Public methylation array data (Illumina EPIC and 450K) were collected from a total of 1 892 colorectal samples. Pairwise differential methylation analyses between tissue types were performed for both array types to "double evidence" differentially methylated probes (DE DMPs). Subsequently, the identified DMPs were filtered on methylation level and used to build a binary logistic regression prediction model. Focusing on the clinically most interesting group (adenoma vs carcinoma), we identified 13 DE DMPs that could effectively discriminate between them (AUC = 0.996). We validated this model in an in-house experimental methylation dataset of 13 adenomas and 9 carcinomas. It reached a sensitivity and specificity of 96% and 95%, respectively, with an overall accuracy of 96%. Our findings raise the possibility that the 13 DE DMPs identified in this study can be used as molecular biomarkers in the clinic.

Conclusions: Our analyses show that methylation biomarkers have the potential to discriminate between normal, precursor and carcinoma tissues of the colorectum. More importantly, we highlight the power of the methylome as a source of markers for discriminating between colorectal adenomas and carcinomas, which currently remains an unmet clinical need.

INTRODUCTION

Colorectal cancer (CRC) is one of the leading causes of cancer-related deaths worldwide. More than 1.9 million incidence cases and over 935 000 deaths were observed worldwide in 2020 [1]. In early stages (I and II), 5-year overall survival chances are at least 90%. Detection of CRC in an advanced stage (IV) reduces survival chances to only 15% [2, 3]. Unfortunately, 75% of all CRC tumors are discovered in advanced stages. Therefore, early detection of colorectal tumors must clearly improve.

CRC is a very heterogeneous disease that typically develops from pre-cancerous lesions, so-called adenomas. In 80% of cases, CRC develops through the adenoma-carcinoma sequence, a process that can take up to 10 years since adenomas tend to progress slowly, with increasing size and dysplasia over time [4, 5]. It has already been demonstrated that both genetic and epigenetic alterations are acquired in the tumor genome during carcinogenesis [5]. Epigenetic alterations have been studied over the past years and have revealed the relation between specific gene expression patterns apart from genetic mutations [5, 6].

One of the most studied epigenetic modifications is DNA methylation. In CRC, widespread hypomethylation blocks have been observed, as well as hypermethylation of specific CpG islands in gene-specific promotors [5, 7]. Despite many efforts, there is still a lot to discover at a molecular level for methylation in colorectal tissue. Particularly, methylation patterns in precancerous colorectal lesions, notably adenomas, are understudied. Online available datasets such as The Cancer Genome Atlas (TCGA) or Gene Expression Omnibus (GEO) mostly include methylation data of invasive tumor tissue. As methylation occurs in very early stages of carcinogenesis, DNA methylation biomarkers are the most compelling candidates for early detection of cancer [5]. Therefore, the DNA methylation data of adenomas are of extreme importance.

In previous research [7], it was already demonstrated that normal tissue and colorectal cancer tissue can be discriminated based on differentially methylated CpG sites. The study was based on publicly available data, which lacks the information on methylation of precancerous lesions as described earlier. Other researchers [5] have investigated differential methylation in normal and low-grade versus high-grade adenomas. Although this study shows very promising results for early biomarker candidates, it lacks a comparison with colorectal cancer tissue. Up until this moment, there is no possibility to discriminate colorectal adenomas from adenocarcinomas with molecular biomarkers in the clinic. However, such biomarkers would be an interesting and important tool for earlier described reasons.

To our knowledge, epigenome-wide analysis of normal, adenoma and colorectal tumor tissue has never been performed simultaneously. Therefore, the goals of this study were: to 1) explore and compare the epigenome of normal colorectal tissue, adenomas and colorectal tumor tissue in one experiment and 2) to identify molecular biomarkers that can discriminate especially between colorectal carcinoma and adenoma. Based on

currently available data, we hypothesized that each of the three tissue types would have a different methylation pattern.

METHODS

Study population, sample collection and pathologist review

A total of 55 samples were requested at the Biobank and the pathology department of the Antwerp University Hospital. Three different types of samples were used: 19 normal adjacent, 17 adenoma and 19 colorectal tumor tissue samples. This included 10 pairs of colorectal cancer and normal samples and 1 pair of adenoma- normal samples of the same patient. Tissue specimens were formalin-fixed paraffin-embedded (FFPE). For each specimen, 10 sections of 6 μm were made of which one slide was stained with hematoxylin and eosin for histologic review. A pathologist verified the tissue type and estimated the percentage of tumor cells. To limit the contamination by non-tumor cells, macrodissection was performed where possible. All samples had at least 50% tumor cells.

DNA extraction and processing

DNA was isolated using the QIAamp FFPE Tissue kit (Qiagen, Hilden, DE) according to the manufacturer's instructions. It is known that FFPE samples generally perform poorly on array-based applications due to the highly degenerated DNA. Therefore, the quality of the DNA was verified using the Infinium FFPE QC kit (Illumina Inc., San Diego, CA, USA) according to the manufacturer's protocol. Only samples with good amplification for all replicates and a maximal ΔCq (difference in quantification cycles compared to the standard) below 5 were selected for use in the bisulfite conversion and restoration step. Bisulfite conversion was performed using the EZ DNA Methylation kit (Zymo Research, Freiburg im Breisgau, DE), according to the manufacturer's instructions. The array-specific incubation program was used for all samples. After bisulfite conversion, DNA samples were restored using the Infinium HD FFPE Restoration kit (Illumina Inc.).

In-house experimental methylation dataset

In total, 55 clinical samples were obtained and processed, the details of which are available in Additional Tables 1 and 2. The Illumina Human MethylEPIC® v1.0 BeadChip (Illumina Inc.) [8] was used to interrogate more than 850 000 CpG sites (probes) genome-wide at single-nucleotide resolution. Raw intensity array data were processed using the minfi (v 1.42.0) R package [9]. Methylation levels were reported as β-values ranging from 0 for unmethylated probes to 1 for fully methylated probes. For quality control, the ratio of log2 median intensities (methylated and unmethylated) along with β-value densities was calculated. β-values were then further preprocessed using ChAMP (v 2.21.1) [10] where probes with a detection p-value > 0.01 in more than

50% of the samples were removed. Control probes, X-/Y chromosome probes, multihit probes, and probes with known single nucleotide polymorphisms (SNPs) were filtered out of the analyses. BMIQ normalization was used to reduce the technical variation of Type-I and Type-II Illumina probes [11]. Out of 55 samples, 28 samples failed quality check and were removed from downstream analyses. The final analyses included 27 samples with 740 330 autosomal probes each (Additional Table 2).

Public methylation datasets

Array data from both Illumina Infinium HumanMethylation450 (more than 450 000 CpG sites) and Human MethylEPIC® BeadChips were downloaded from several public data repositories including GEO, TCGA and the Array Express databases. A total of 1 116 450K and 786 EPIC samples were acquired, the accession numbers and full details of which can be found in Additional Table 3. To ensure consistent data processing, we opted to use signal intensity or raw idat files. The datasets were then processed using the same steps described above for the in-house experimental methylation data. Out of the total 1 879 samples, 14 failed quality check and were removed from downstream analyses.

Ethical approval

The study was conducted under Good Clinical Practice guidelines and the Declaration of Helsinki. Samples used in this study were previously collected in the Biobank of the Antwerp University Hospital and retrospectively used in this study. Patients give consent for the use of their bodily material in research when consenting to an invasive procedure (according to article 20 of the Belgian Law on the procurement and use of human corporal material intended for human application or scientific research of 19 December 2008). Approval for the study protocol (and any modifications thereof) was obtained from the ethical committee of the Antwerp University Hospital (Ref. N°20/02/010). Other data used in this study are publicly available. As such, neither patient consent nor institutional review board approval was required.

Definitions of genomic regions and differential methylation

Genomic region annotations were based on Illumina 450K and EPIC array manifest files and were divided into two main groups. The first consists of genomic locations concerning genes. These included: 1st exon; 3′ UTR (3′ untranslated region), 5′ UTR (5′ untranslated region), Body (gene body), IGR (intergenic regions), TSS1500 (200 to 1500 nucleotides, upstream of the transcription start site, TSS), TSS200 (up to 200 nucleotides upstream of TSS), and ExonBnd (exon boundaries). The second describes annotations of probe location relative to CpG islands. These included: Islands, North shelf (2–4 kb upstream of CpG island), North shore (0–2 kb upstream of CpG island), Open Sea (non-CpG island-related sites), South shelf (2–4 kb downstream of CpG island), and South shore (0–2 kb downstream of CpG island).

Genome-wide DNA methylation was investigated in the context of differentially methylated probes (DMPs), regions (DMRs) and blocks (DMBs). DMPs were defined as CpG sites with statistically significant differences in methylation levels between groups. In contrast, DMRs and DMBs are larger genomic regions— between ~ 10 bp—10 kb and 10 kb—1 Mb, respectively— exhibiting a quantifiable difference in methylation between groups and containing hundreds of CpG sites.

Differential methylation analyses

Differential methylation analysis was carried out via ChAMP (v 2.21.1), which uses parametric linear mixed models to test differences in methylation between groups [10]. A two-level, three-way differential methylation analysis was performed in the public EPIC datasets; adenoma versus normal tissue, carcinoma versus normal tissue, and adenoma versus carcinoma (Fig. 1). DMP p-values were adjusted for multiple testing using the Benjamini– Hochberg correction. DMRs and DMBs were identified using an implemented extension of the Bumphunter algorithm in ChAMP, with minimum sizes of 50 and 500 bp, respectively. Gene set enrichment analysis (GSEA) was done using the ChAMP and methylGSA R packages [12]. Differential methylation analysis was carried out on the public methylation datasets which constituted the discovery cohort (Fig. 1).

Double evidencing and biomarker selection

DMPs that were first identified through the EPIC differential methylation analyses and that were later also found in the 450K analyses were termed double evidenced DMPs (DE DMPs). These DE DMPs represent important methylation markers that are identified through the analysis of two separate, large datasets. The criteria for assigning DE DMPs for use in the sub- sequent models were 1) a $|\Delta\beta| \geqslant 0.3$ [13] and 2) a corrected p-value $\leqslant 0.01$ in both array types. After merging DMP lists and screening for DE DMPs, binary logistic regression models were fitted to predict tissue type (adenoma/carcinoma/normal tissue) using the specific DE DMPs for each of the three analyses (Fig. 1). To test over-fitting, all models were tenfold cross-validated. Prediction metrics were assessed by plotting receiver operating characteristic (ROC) curves, and confusion matrices were also generated to calculate overall sensitivity, sensitivity and accuracy. The final model was then validated in the in-house experimental methylation datasets which constituted the validation cohort. Prediction metrics were also calculated for the validation model.

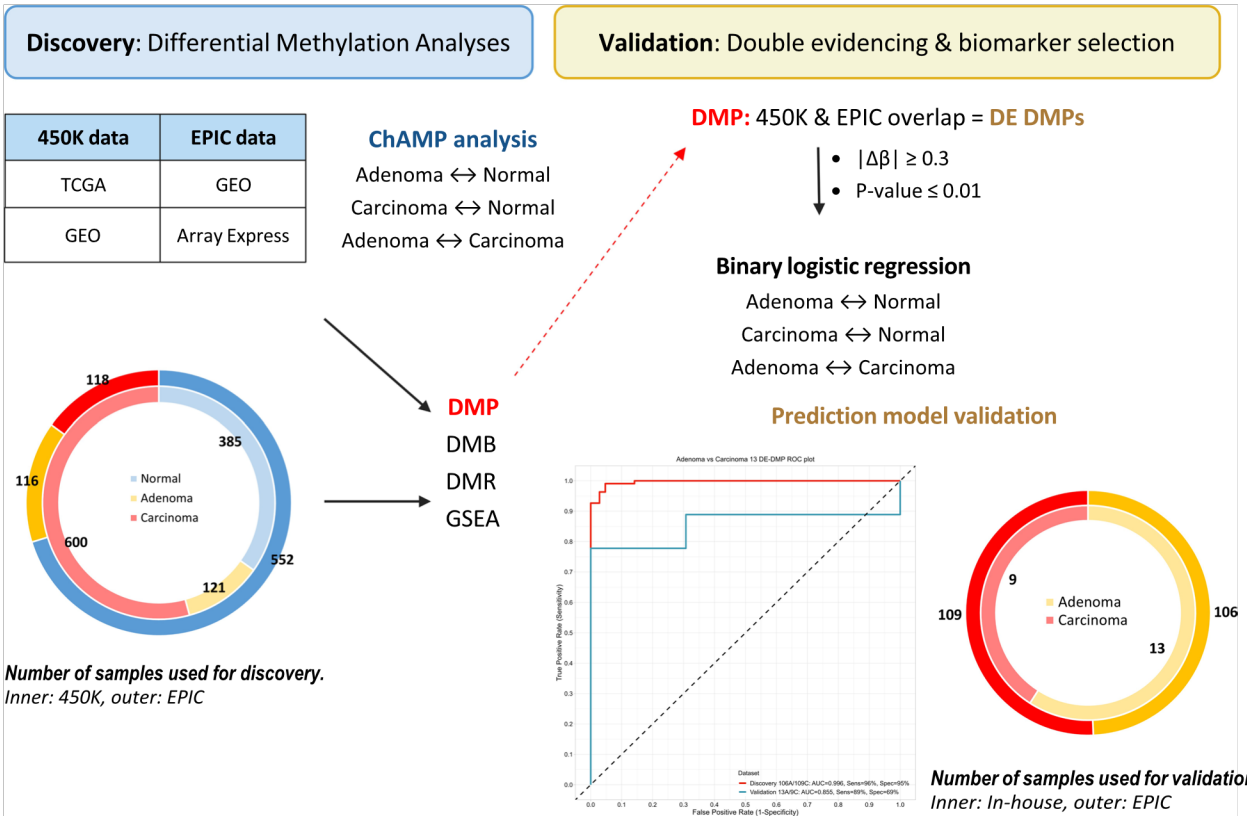


Figure 1 | Study overview. DMP= differentially methylated probe, DMB= differentially methylated block, DMR= differentially methylated region, GSEA= gene set enrichment analysis, DE= double evidenced

Statistical analyses

The statistical software R (v 4.2.0) [14] was used for all analyses and visualizations. In all regression models, age was accounted for as a covariate, but was excluded from the final model if its effect on the outcome was not significant. Unless stated otherwise, all reported p-values are two-sided, and those ⩽ 0.01 were considered statistically significant. All genomic annotations were based on the GRCh37/hg19 genome build.

RESULTS

Genome-wide methylation profiling

To comprehensively explore the difference in methylation patterns between normal, adenoma and carcinoma tissue, DNA methylation was profiled pairwise between the three tissue types. This genome-wide differential methylation profiling was carried out on public EPIC array datasets. The results of these analyses are summarized in Table 1. Sizeable genome-wide DNA methylation differences were observed between the three tissue types (Fig. 2). β-values in all three tissues exhibited characteristic bimodal distributions (Fig. 2A), while

on average normal tissues had the highest methylation levels followed by adenomas and lastly carcinomas (Fig. 2B). Based on the widespread differences in methylation, the three tissues clustered independently using both multidimensional scaling (MDS) and t-distributed stochastic neighbor embedding (tSNE) approaches (Fig. 2C, D). MDS is used for the visualization of outliers, while tSNE rather shows how samples group together. In our analyses, both methods agreed. The tSNE plot shows four distinct clusters for normal tissue (N). The tissues formed mostly discernable clusters where (pre)malignant lesions (i.e., adenomas (A) and carcinomas (C)) could be clearly resolved from N. However, A and C clustered more closely together (Fig. 2C, D).

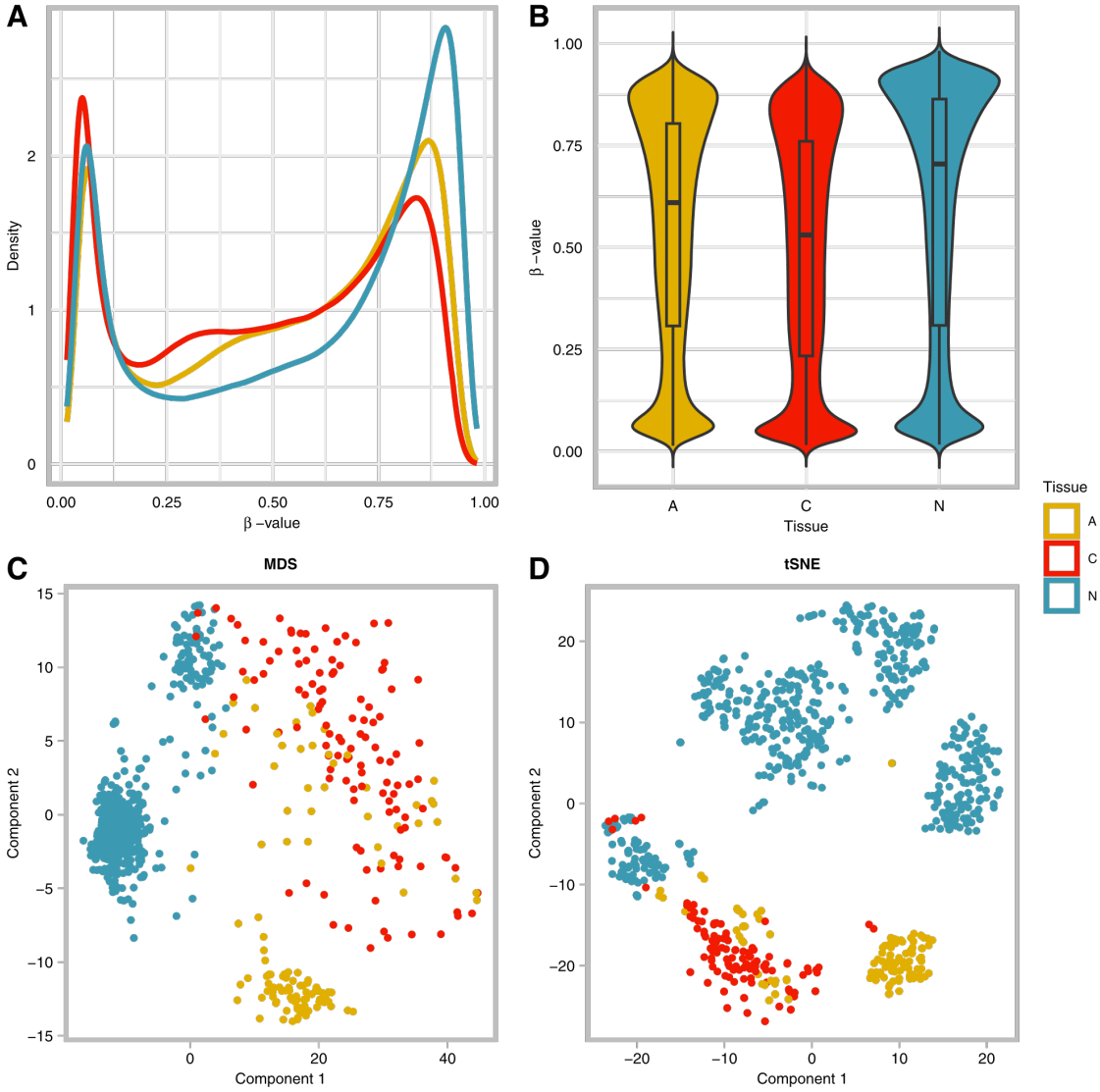


Figure 2 | Landscape of DNA methylation of adenoma, carcinoma, and normal colorectal tissues in EPIC datasets. **A.** Density plot showing the characteristic bimodal distribution of methylation β-values in all 3 tissues based on EPIC array data. **B.** Violin plot of the mean methylation in each of the tissues, shows overall methylation decreases with increase in malignancy. **C.** MDS plot highlighting the data structure and sample relationship among the tissue groups in EPIC array data. **D.** tSNE plot showing a defined cluster for each of the different tissues, highlighting the ability to resolve samples based on their methylation patterns, despite overlap between adenomas and carcinomas. MDS= multidimensional scaling, tSNE= t-distributed stochastic neighbor embedding, A= adenoma, C= carcinoma, N= normal tissue

Table 1 | Summary of DMPs, DMRs and DMBs in all three analyses

	Comparison	Adenoma versus normal	Carcinoma versus normal	Adenoma versus carcinoma
DMB	EPIC	703	582	1552
DMR	EPIC	3510	6756	5067
DMP	450K	344,165	304,548	170,300
	EPIC	620,643	693,813	558,897
	Common DMP (450K and EPIC)	257,141	258,853	124,082
	DE DMP with $\lvert \Delta\beta \rvert \geq 0.3$ in EPIC AND 450K	62	56	**13**

Bold value indicates p value ⩽ 0.01

DMPs

When studying differences in DNA methylation at a single-base resolution, we identified 620 643 DMPs in A vs C. When C vs N was compared, 693 813 DMPs were observed while comparing A vs N resulted in 558 897 DMPs (see EPIC data in Table 1). The distribution and location of these DMPs in relation to genomic features and CpG islands are shown in Fig. 3A. In each com- parison, most DMPs were in the gene body (36.84% on average), which is expected based on the distribution of probes on the EPIC array [15]. This was followed by the intergenomic regions (28.16% on average) and TSS1500 (12.66% on average). We also found DMPs located in the 5'UTR (8.53% on average), in the TSS200 (7.64% on aver- age), the 1st exon (3.04% on average) and 3'UTR (2.44% on average). Lastly, the exon boundaries were studied, but they only represented 0.68% of DMPs (Fig. 3A). Concerning DMP location in relation to CpG islands, the largest proportion of DMPs mapped to open-sea regions (55.66% on average) followed by CpG islands (19.13% on average). North shores contained ± 9.92% of DMPs, while south shores contained on average 8.47% of DMPs. North and south shelves contained the lowest average proportion of DMPs at 3.53% and 3.28%, respectively (Fig. 3A). Definitions of DMP locations can be found in the materials and methods section.

The majority of DMPs were hypomethylated compared to hypermethylated (80.67% in A vs N, 94.21% in C vs N and 91.08% in A vs C, in Fig. 3B). When evaluating the tissue types, most DMPs were hypomethylated in tissue types with a higher degree of malignancy (given that the malignant potential increases from normal, to adenoma and eventually carcinoma) (Fig. 3B). To allow for a com- parison between the three tissue types, the DMP counts are normalized by dividing them through the total number of analyzed CpG sites in each category.

DMRs and DMBs

To study small regions with differential methylation that might be functionally involved in transcriptional regulation, DMRs between the three tissue types were studied. Most DMRs were identified when C vs N were compared, followed by the comparison of A vs C and the smallest number of DMRs were identified when A vs N were compared (6 756, 5 076 and 3 510 DMRs, respectively) (Fig. 3C). Since it has been reported that large

hypomethylated blocks are a universal feature of cancer tissue, methylation data was analyzed to identify DMBs for the comparison of the three tissue types. We identified 1 552 DMBs when comparing A vs C, 703 DMBs when comparing A vs N and lastly 582 DMBs for C vs N (Fig. 3C). Definitions of DMRs and DMBs can be found in the materials and methods section “Definitions of genomic regions and differential methylation”.

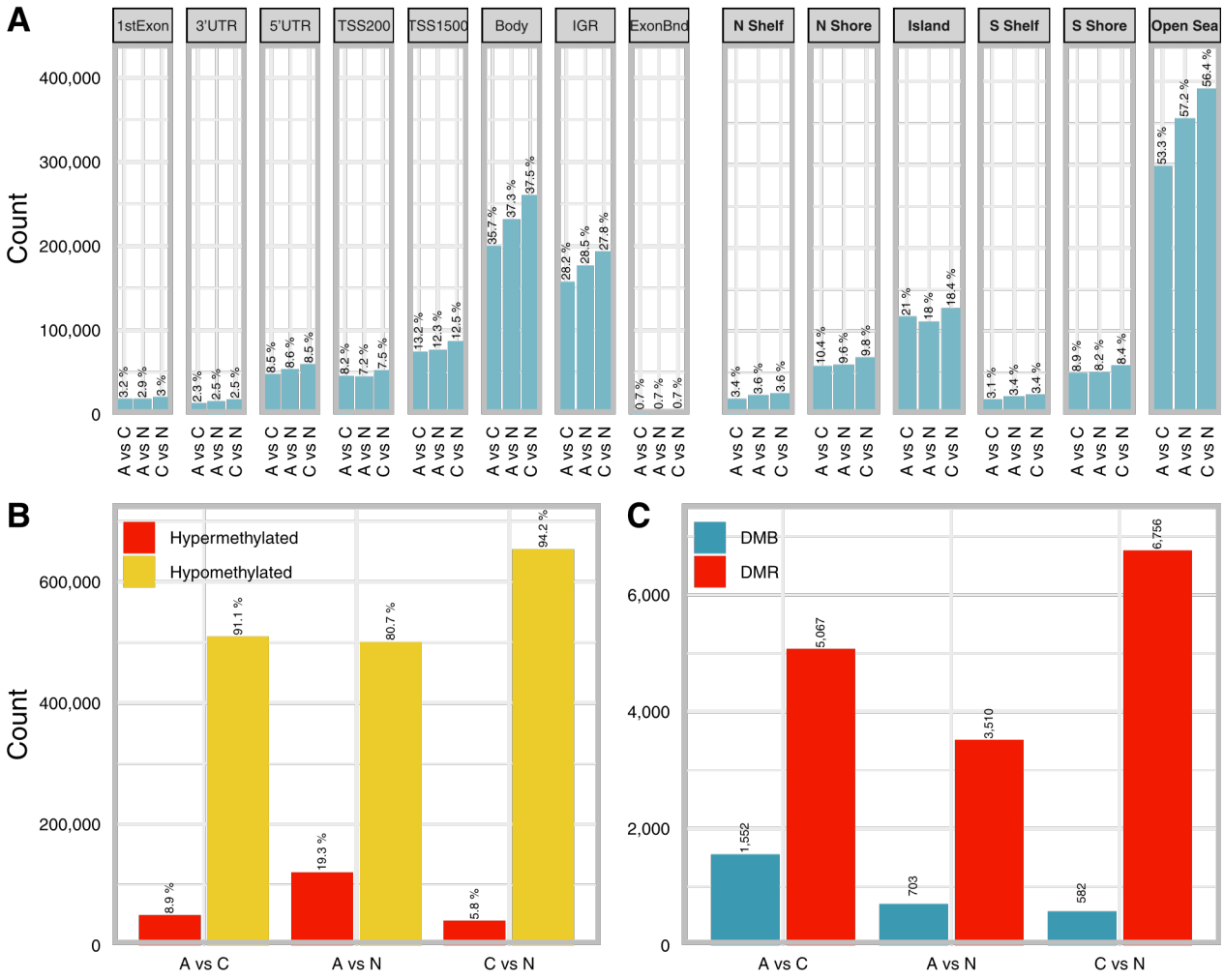


Figure 3 | Differential methylation in adenoma, carcinoma, and normal colorectal tissues in both EPIC and 450K datasets. **A.** Barplot of DMP counts per genomic region for each of the 3 pairwise comparisons for both methylation platforms. Percentages are are fractions of the total DMPs for each comparison and platform. **B.** Barplot of hyper- ($\beta \geq 0.7$) and hypomethylated ($\beta \leq 0.3$) DMPs for each of the 3 comparisons for both methylation platforms. Percentages are fractions of the total DMPs for each comparison and platform. **C.** Barplot of DMB and DMR counts for all 3 comparisons for both methylation platforms. Annotations in regular font are with reference to genes, those in bold are with reference to CpG islands. UTR= untranslated region, IGR= intergenic region, TSS= transcription start site, N= north, i.e., upstream (5') of CpG island, S= south, i.e., downstream (3') of CpG island, ExonBnd=exon boundaries, DMB= differentially methylated block, DMR= differentially methylated region

Double evidenced differential methylation (DE DMPs)

To double evidence the DMPs identified through the public MethylEPIC® dataset, analysis of additional Illumina 450K data of 385 normal, 121 adenoma and 600 carcinoma samples from public datasets was performed (1 106 450K samples mentioned in Fig. 1). The common DMPs that were detected in the datasets of both the EPIC and 450K methylation arrays and had an absolute delta beta value of > 0.3, were termed double evidenced DMPs (DE DMPs). Additional Fig. 1 represents an overview of the unique and common DMPs in the three different tissue groups. Sixty-two DE DMPs were identified when comparing adenoma and normal tissue, 56 DE DMPs for carcinoma and normal tissue and 13 DE DMPs for adenoma and carcinoma tissue (shaded row in Table 1). More information regarding the location of the DE DMPs within the genome can be found in Additional Table 4.

Methylation as a biomarker for adenomas and carcinomas

To test the discriminatory power of methylation markers in classifying adenomas versus carcinomas, which are the most difficult to resolve clinically, a binary logistic regression model was built using the 13 DE DMPs reported above as predictors. Clustering both the public data (Fig. 4A–C) and the in-house data (Fig. 4D) using the 13 DE DMPs resulted in distinct clusters between adenomas and carcinomas and more unified groupings than using the array data as a whole. Hierarchical clustering revealed that these DMPs were more hypermethylated in adenomas and hypomethylated in carcinomas (Fig. 4A). Clustering the public data could clearly resolve the 2 tissue types, albeit some samples remained doubtful (Fig. 4B, C). Clustering the in-house data fared better, resulting in 2 separate clusters with only 2 of the carcinomas localizing in the adenoma cluster (Fig. 4D).

The final model was trained on the public EPIC array data and validated in the in-house experimental methylation datasets (Fig. 1, methods). Importantly, the classifier model reached a cross-vali- dated area under the curve (AUC) of 0.996 and 0.855 in the discovery and validation datasets, respectively. Sensitivities and specificities at different cut-off values for the predicted probabilities are shown through a ROC plot (Fig. 4E). At optimal cut-off, a sensitivity of 96.33% and a specificity of 95.28% for the detection of carcinomas versus adenomas were reached, with an overall accuracy of 95.81% and a misclassification error rate of 4.19%. In the in-house data, the model successfully classified 13 out of 13 adenomas and 7 out of 9 carcinomas. In all, the model exhibited high predictive power and good generalizability across different datasets. The results of the validation of the DE DMPs for comparison of adenoma vs normal and carcinoma vs normal are reported in Additional Fig. 2 and Fig. 3. In addition, a circos plot representing the genome-wide differential methylation between adenoma and carcinoma tissue is provided in Fig. 5. This plot depicts the DMPs, DMRs and DMBs of A vs C in view of the epigenome and compared to known CRC biomarkers.

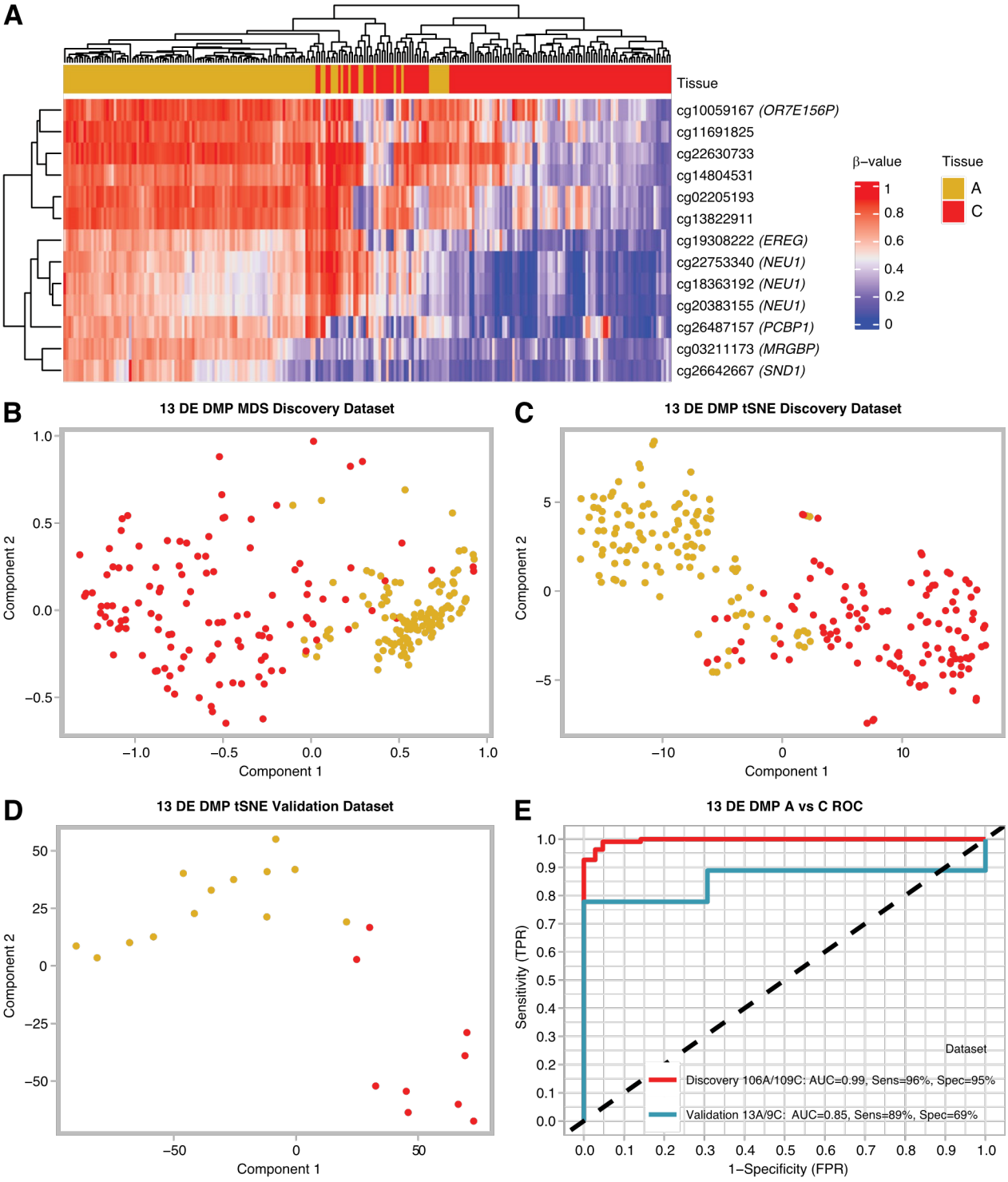


Figure 4 | The selected 13 DE DMP markers were effective at classifying adenomas and carcinomas. **A.** Heat map and hierarchical clustering analysis of the discovery EPIC dataset based on the 13 identified DE DMP markers shows a block like structure with almost half of the markers being hypermethylated in carcinoma and hypomethylated in adenomas and vice versa for the other half. **B.** MDS clustering of the discovery dataset using the 13 markers shows 2 distinct clusters. **C.** tSNE clustering of the discovery dataset using the 13 markers could also resolve the two tumor types. **D.** tSNE clustering of the validation dataset using the 13 markers shows a clear separation between adenomas and carcinomas, only 2 carcinomas are falsely classified. **E.** ROC curves for the final 13 DE DMP classifier model for both discovery and validation datasets from

EPIC arrays. Sensitivity and specificity, for distinguishing between adenomas and carcinomas, at various cut-off values for the datasets are plotted. The model yielded an AUC of 0.99 and reached a sensitivity and specificity of 96.33% and 95.28%, respectively, while overall model accuracy was 95.81% in the discovery dataset. In the validation dataset it had an AUC of 0.85, and reached a sensitivity and specificity of 89.36% and 69.78%, respectively. The diagonal dotted line represents the line of no discrimination between the two tumor types. DE DMP double evidenced differentially methylated probes, ROC receiver operating characteristic, MDS multidimensional scaling, tSNE t-distributed stochastic neighbor embedding, TPR = true positive rate, FPR = false positive rate, A = adenoma, C= carcinoma

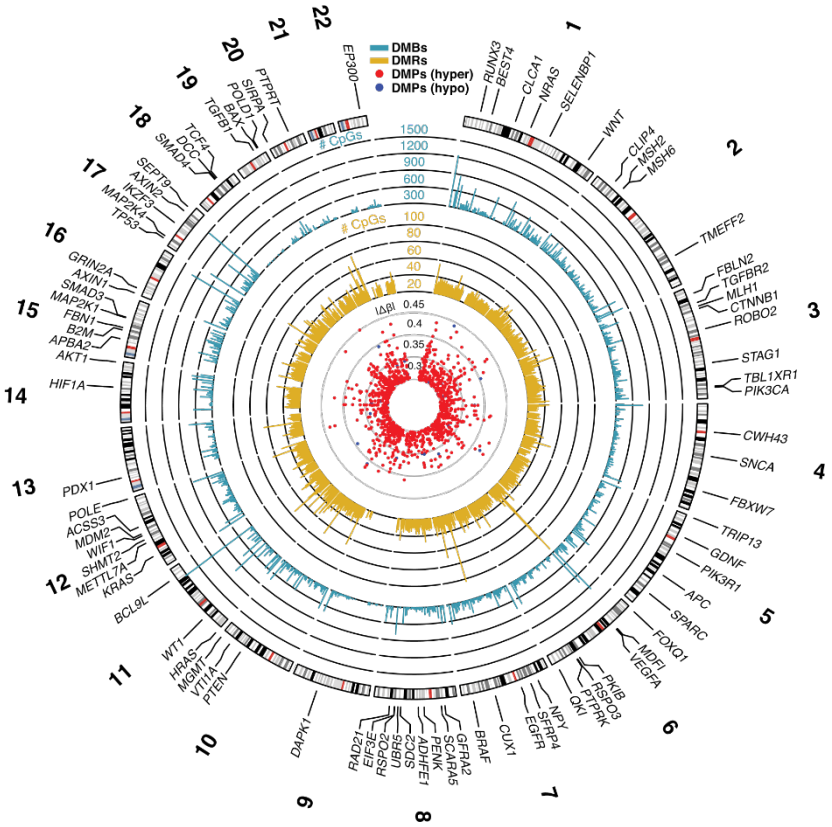


Figure 5 | Circular genome plot summarizing genome-wide differential methylation between colorectal adenoma and carcinoma tissue. The of DMPs represents their |Δβ| in carcinomas. DMBs= differentially methylated block, DMRs= differentially methylated regions, DMPs= differentially methylated probes, hyper= hypermethylated, hypo= hypomethylated

Gene set enrichment analysis

Reactome, gene ontology (GO) and KEGG functional enrichment analysis were performed to better understand the functional implication of differential methylation patterns in adenoma vs carcinoma samples. Pathways were selected based on p-values, with a cut- off at < 0.01. We identified 1 375, 111 and 32 pathways for GO, Reactome and KEGG analyses in A vs C, respectively (Additional Table 5). The top 10 most significantly enriched gene sets in each category are represented in Additional Fig. 4. Functional terms that were highly enriched included terms related to developmental pathways, cell organelles, metabolism, signaling and response mechanisms. Searching for overlapping pathways within the three analyses, the MAPK signaling, cell cycle,

ubiquitin-mediated proteolysis, endocytosis and Wnt signaling pathway were found to be significantly enriched. Furthermore, genes within the enriched pathways were investigated in more detail. The *NEU1* gene, which contained 3 DE DMPs for A vs C, was found in all three GSEAs. Pathways including the *NEU1* gene are mentioned in Additional Table 6. All frequently mutated genes in cancer are registered in the Catalogue of Somatic Mutations in Cancer (COSMIC) database. This list of genes provides valuable insights into the genetic mechanisms underlying cancer. Within the COSMIC genes, 173 genes were found that were present in all three GSEAs (Additional Table 7). For example, *BRAF*, *HRAS*, *MLH1* and *EGFR* were found to be enriched in the GSEA.

DISCUSSION AND CONCLUSIONS

Previous research has demonstrated the methylome's potential for the discovery of biomarkers. In CRC, it has been shown that normal and colorectal cancer tissue, as well as low-grade and high-grade adenomas, can be discriminated based on methylation pattern [5–7]. Therefore, we performed the first study to explore and compare the epigenome of normal colorectal tissue, precancerous lesions (adenomas) and colorectal cancer tissue together and to identify biomarkers that can discriminate between these three tissue types. Based on the current available literature, we hypothesized that each of the three tissue types would be differentially methylated.

Our results are consistent with this hypothesis. We identified numerous DMPs, DMBs and DMRs between the three tissue types (Table 1). The most interesting aspect is that when normal colorectal tissue is compared to adenoma or carcinoma tissue, most of the DMPs were hypomethylated in the tissue type with increasing malignant potential (Fig. 3A, B), which indicates an important role for hypomethylation in carcinogenesis. This is in accordance with previous studies that indicated widespread hypomethylation in cancer tissue compared to healthy tissue, which is observed across cancer types [16, 17]. It also corresponds to the findings of Fan *et al.*, who observed increasing DNA hypomethylation starting from low-grade adenoma stage, leading to further hypomethylation at high-grade adenoma and CRC stage [5]. Likewise, Liu *et al.* found significantly more hypomethylated DMPs than hypermethylated DMPs in adenoma tissue compared to adjacent normal tissue. For DMRs, the same pattern was observed [18]. When focusing on the difference in methylation between the three tissue types, it is interesting to note that not all normal samples were alike. In the MDS and tSNE plots (Fig. 2C, D), two and four distinct subclusters for the normal samples can be observed, respectively. This indicates the possibility of several subtypes of normal colon tissue with different methylation patterns. We observed different clusters based on sample location (left vs right, data not shown), which has also been described in literature before [19–22]. However, healthy colon tissue adjacent to the tumor tissue was used instead of normal colon samples of healthy patients. In literature, the phenomenon of field cancerization has been described, where amongst others epigenetic changes have been reported in normal colon mucosa adjacent to the tumor [23–

27]. Hawthorn *et al*. described chromosomal instability in regions surrounding the tumor as far as 10 cm distal [23]. Park *et al*. described the aberrant methylation of non-adjacent normal-appearing tissue [25]. Unfortunately, for most of the datasets, there is no information on the distance at which the normal-looking tissue was taken, making it difficult to estimate the field effect. However, this clinical information is also lacking in public data. Lastly, two distinct morphological pathways of CRC carcinogenesis exist, potentially explaining the two clusters found in the MDS plot (Fig. 2C). Both the conventional and the alternative/serrated pathways are characterized by specific epigenetic alterations. Different mechanisms lay behind these pathways, which are associated with MSI status and CpG island methylator phenotype (CIMP). A specific CRC classification of five molecular subtypes based on MSI and CIMP status has been described previously. The four distinct clusters found in the tSNE plot (Fig. 2D) could potentially be explained by these molecular subtypes, but this cannot be verified due to the lack of clinical data [28].

When comparing our DMPs to those found in literature, we find many similarities. For example, CpGs in the *ADHFE1* [5], *SND1*, *OPLAH*, *TMEM240*, *NR5A2*, *TLX2*, *COL4A1*, *ZFP64* [13], *MYO1G* [29], *CREB1* [18], *NPY* and *PENK* [30] genes were also identified in other studies comparing the methylation pattern of healthy colorectal tissue to adenoma and/or carcinoma tissue. Several of these methylation markers can also be appreciated from the circos plot (Fig. 5).

From a clinical perspective, the difference between colorectal adenoma and carcinoma is the most relevant. Therefore, a more in-depth analysis was performed on the difference in methylation between those 2 tissues (Fig. 4). When comparing their methylation patterns, surprisingly 3 out of 13 DE DMPs were located on chromosome six. Chromosome six is a well-known chromosome in oncology. It contains several clinically important proto-oncogenes as well as the major histocompatibility complex. Several genes linked to CRC are located on this chromosome, including *ROS1*, *VEGFA*, *CDKN1A* and *VIP*. A total of 37 797 DMPs was found in the EPIC analysis. 7 810 thereof were in the major histocompatibility complex (MHC). The MHC contains more than 100 genes implicated in the immune response, including HLA-A, HLA-B and HLA-C. These genes and the MHC class I molecules play an important role in the anti-cancer immune response [31]. Downregulation of MHC class I has been observed in 40–90% of cancer types and was often correlated with a worse prognosis [32].

The three DE DMPs on chromosome 6 were located in the TSS1500 (shore) region of the neuraminidase 1 (*NEU1*) gene. This gene encodes a protein that functions as a lysosomal enzyme. It cleaves terminal sialic acid residues from its substrates including glycoproteins/glycolipids. It has no clear cancer-related function, but it is described to play a role in other pathways for the innate immune system, glycosphingolipid metabolism, diseases of glycosylation and synthesis of substrates in N-glycan biosynthesis [33], which were also found to be enriched in the GSEAs (see Additional Table 6). Furthermore, three publications have already described a link between *NEU1* and CRC. In 2009, Uemura *et al*. reported the regulatory role of *NEU1* in integrin β4-mediated

signaling, which led to the suppression of metastasis [34]. Almost a decade later, Forcella *et al*. found that human sialidases are severely dysregulated in several tumors and described their potential application in cancer diagnosis [35]. Jiao *et al*. further underlines the role of *NEU1* in tumorigenesis regulation through several pathways, including immune-mediated tumorigenesis and regulation of vascularization [36]. In addition, two other DE DMPs reported in this study (Fig. 4), are reported in previous methylation studies in CRC. First, *EREG* methylation and subsequent low *EREG* gene expression were correlated with poor response to anti- *EGFR* therapy in colorectal cancer [19, 37, 38]. Further- more, *SND1* methylation was identified as one of the top 14 methylation markers for discriminating between CRC and normal tissue in a study by Naumov *et al*. [13].

Later, the 13 DE DMPs detected through the comparison of the methylation pattern of colorectal adenoma and carcinoma were used to build a model that can discriminate between these two lesions (Fig. 4E). These 13 DE DMPs represent the most significant differences between these two tissue types. During the validation of the prediction model in the in-house experimental methylation dataset, an increased error rate was noted (from 4.19% to 11.62%). This might be due to the smaller group of samples in the validation group and/ or due to a lower quality of methylation arrays run on FFPE tissue instead of fresh frozen tissue.

When comparing the performance of our model to other methylation models, it is interesting to compare to *SEPT9*. This is the best-known example of DNA methylation as a biomarker in CRC and was commercialized as the EpiProColon® assay. Although the use of this assay has proven effective for CRC detection, it lacks sensitivity for the detection of adenomas. Sensitivities ranging from 11.2% to 31.8% for methylated *SEPT9* in adenomas have been reported [39]. Combinations with other markers, for example *ALX4*, increased the sensitivity to 37%, which shows there is plenty of room for improvement [40]. Our model, combining 13 DE DMPs, yielded a sensitivity of 96% for discriminating adenomas and carcinomas. All 13 adenomas were correctly classified. This is already a major improvement compared to methylated *SEPT9*, although more research and external validation will be needed to prove the superiority of the 13 DE DMPs.

An aspect of working with public data is the lack of quality control. When the data of publicly available methylation array data were analyzed for this study, certain samples included in these datasets were not able to pass quality control and had to be excluded. Therefore, it is advised to download the signal intensity or raw idat files and not β-values, to perform the quality control yourself to ensure adequate quality.

One of the limitations of this study is the use of FFPE material for methylation arrays. A known problem when FFPE samples are used for methylation arrays is the fact that this often results in lower quality data. Previous studies showed that a restore method can result in reliable and high-quality epigenomic data, concordant to that of fresh frozen tissue [41–43]. Therefore, the Infinium HD FFPE Restoration kit was used in this study. However, in our analyses it was noticed that the results of the vali- dation of the 13 DE DMPs were sample dependent. The sample age did not affect the quality of FFPE-derived DNA, which is in concordance with the

study of Kling *et al.* [44]. When the model was tested with lower-quality samples (without the expected bimodal distribution of beta values and more beta values around 0.5), it per- formed worse (data not shown) [44]. High-quality data are thus needed for reliable analyses. Even though different pretreatment processes were developed to reduce formalin artifacts [41], restoring FFPE samples was not found to be effective in our study. This resulted in a limited number of samples used in this study.

Another restriction of this study is that only one out of 13 adenomas from the in-house experimental methylation dataset was high-grade, while all others were low-grade. It would be of interest to identify methylation markers to make a distinction between low-grade and high-grade adenomas, since this might allow for minimally invasive identification of high-grade adenomas, which are known to have a higher risk of developing into carcinoma. However, since only one of our adenoma samples was high-grade and that dysplasia grade was not reported for most adenomas in the public datasets, this analysis was not possible. However, this comparison has been reported by Fan *et al.* [5].

Due to the stability of DNA methylation and the fact that aberrant methylation occurs early in carcinogenesis, the methylome has been considered an ideal source for potential biomarkers. The findings of this study raise the possibility that the 13 DE DMPs identified in this study can be used as targets for a liquid biopsy assay to distinguish adenoma from carcinoma in a minimally invasive way. The non-invasive detection of colorectal adenoma and carcinoma and the distinction between these lesions is highly clinically relevant. Early detection and removal of these lesions in the colorectum can prevent the development and locoregional or metastatic spread of colorectal cancer. Most adenomas and carcinomas are detected through colorectal cancer screening with fecal occult blood tests and subsequent colonoscopy. However, for certain patient groups these tests are not ideal, and a minimally invasive test is preferred. For example, in patients with congestive heart disease the fluid load of bowel preparation should be avoided and in patients who are treated with anticoagulants an invasive colonoscopy with biopsy for histopathological analysis can cause bleeding. Since only a small proportion (± 5%) of adenomas will eventually progress to carcinoma and this process takes up to 5–10 years, the removal of an adenoma is less urgent than the removal of a carcinoma. Therefore, it is of clinical importance to not only detect these lesions minimally invasively, but also to discriminate between these two tissue types, since treatment and follow-up will be different. In addition, a minimally invasive method to do this (e.g., liquid biopsy or stool samples with the 13 DE DMP markers), would be an important added value. With this study, we demonstrate the strength of differentially methylated CpG sites to be used in the clinic as biomarkers. In conclusion, our analyses highlight the power of the methylome, showing that methylation biomarkers can be used to identify colorectal adenoma and carcinoma, but also have the potential to discriminate between these two tissue types.

DECLARATIONS

Acknowledgements

The authors would like to thank all patients and the UZA biobank for their contribution to this study.

Funding

Research in the lab of Prof. Guy Van Camp is supported by grants awarded by the University of Antwerp (IOF/SBO 43782, BOF/Methusalem grant 40790, BOF/TOP 39705). Joe Ibrahim and Isabelle Neefs are supported by a strategic basic PhD fellowship of the Research Foundation Flanders (FWO; 11B5220N and IS67523N).

Data availability

The dataset supporting the conclusions of this article is available in the EGA European Genome-Phenome Archive (accession number: EGAS00001007017).

Competing interests

The authors declare no competing financial interests.

REFERENCES

1. GLOBOCAN I. Cancer Today: WHO; 2020 [Available from: https:// gco. iarc. fr/ today/ online- analy sis- multi-bars.
2. Cervantes A, Adam R, Rosello S, Arnold D, Normanno N, Taieb J, et al. Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up(dagger). Ann Oncol. 2022.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7–34.
4. Liu C, Fennell LJ, Bettington ML, Walker NI, Dwine J, Leggett BA, et al. DNA methylation changes that precede onset of dysplasia in advanced sessile serrated adenomas. Clin Epigenet. 2019;11(1):90.
5. Fan J, Li J, Guo S, Tao C, Zhang H, Wang W, et al. Genome-wide DNA methylation profiles of low- and high-grade adenoma reveals potential biomarkers for early detection of colorectal carcinoma. Clin Epigenet. 2020;12(1):56.
6. Jung G, Hernandez-Illan E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. Nat Rev Gastroenterol Hepatol. 2020;17(2):111–30.
7. Ibrahim J, Op de Beeck K, Fransen E, Croes L, Beyens M, Suls A, et al. Methylation analysis of Gasdermin E shows great promise as a biomarker for colorectal cancer. Cancer Med. 2019;8(5):2133–45.
8. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8(3):389–99.
9. Wang Z, Wu X, Wang Y. A framework for analyzing DNA methylation data from Illumina Infinium HumanMethylation450 BeadChip. BMC Bioinf. 2018;19(Suppl 5):115.
10. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics. 2017;33(24):3982–4.
11. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, GomezCabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189–96.
12. Ren X, Kuan PF. methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing. Bioinformatics. 2019;35(11):1958–9.
13. Naumov VA, Generozov EV, Zaharjevskaya NB, Matushkina DS, Larin AK, Chernyshov SV, et al. Genome-scale analysis of DNA methylation in colorectal cancer using Infinium HumanMethylation450 BeadChips. Epigenetics. 2013;8(9):921–34.
14. Team RC. R: a language and environment for statistical computing. 2017.
15. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016;17(1):208.

16. Ehrlich M. DNA methylation in cancer: too much, but also too little. Oncogene. 2002;21(35):5400–13.
17. Timp W, Bravo HC, McDonald OG, Goggins M, Umbricht C, Zeiger M, et al. Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. Genome Med. 2014;6(8):61.
18. Liu L, Chen Y, Liu T, Yu J, Ma L, Wu H. Genome-wide DNA methylation profiling and gut flora analysis in intestinal polyps patients. Eur J Gastroenterol Hepatol. 2021;33(8):1071–81.
19. Lee MS, McGuffey EJ, Morris JS, Manyam G, Baladandayuthapani V, Wei W, et al. Association of CpG island methylator phenotype and EREG/ AREG methylation and expression in colorectal cancer. Br J Cancer. 2016;114(12):1352–61.
20. Lee MS, Menter DG, Kopetz S. Right versus left colon cancer biology: integrating the consensus molecular subtypes. J Natl Compr Cancer Netw. 2017;15(3):411–9.
21. Loree JM, Pereira AAL, Lam M, Willauer AN, Raghav K, Dasari A, et al. Classifying colorectal cancer by tumor location rather than sidedness highlights a continuum in mutation profiles and consensus molecular subtypes. Clin Cancer Res. 2018;24(5):1062–72.
22. Sugai T, Habano W, Jiao YF, Tsukahara M, Takeda Y, Otsuka K, et al. Analysis of molecular alterations in left- and right-sided colorectal carcinomas reveals distinct pathways of carcinogenesis: proposal for new molecular profile of colorectal carcinomas. J Mol Diagn. 2006;8(2):193–201.
23. Hawthorn L, Lan L, Mojica W. Evidence for field effect cancerization in colorectal cancer. Genomics. 2014;103(2–3):211–21.
24. Alonso S, Dai Y, Yamashita K, Horiuchi S, Dai T, Matsunaga A, et al. Methylation of MGMT and ADAMTS14 in normal colon mucosa: biomarkers of a field defect for cancerization preferentially targeting elder AfricanAmericans. Oncotarget. 2015;6(5):3420–31.
25. Park SK, Song CS, Yang HJ, Jung YS, Choi KY, Koo DH, et al. Field Cancerization in sporadic colon cancer. Gut Liver. 2016;10(5):773–80.
26. Cherkezyan L, Stypula-Cyrus Y, Subramanian H, White C, Dela Cruz M, Wali RK, et al. Nanoscale changes in chromatin organization represent the initial steps of tumorigenesis: a transmission electron microscopy study. BMC Cancer. 2014;14:189.
27. Guo H, Zeng W, Feng L, Yu X, Li P, Zhang K, et al. Integrated transcriptomic analysis of distance-related field cancerization in rectal cancer patients. Oncotarget. 2017;8(37):61107–17.
28. De Palma FDE, D'Argenio V, Pol J, Kroemer G, Maiuri MC, Salvatore F. The molecular hallmarks of the serrated pathway in colorectal cancer. Cancers (Basel). 2019;11(7).
29. Luo H, Zhao Q, Wei W, Zheng L, Yi S, Li G, et al. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. Sci Transl Med. 2020;12(524).
30. Roperch JP, Incitti R, Forbin S, Bard F, Mansour H, Mesli F, et al. Aberrant methylation of NPY, PENK, and WIF1 as a promising marker for bloodbased diagnosis of colorectal cancer. BMC Cancer. 2013;13:566.

31. Garrido F, Aptsiauri N. Cancer immune escape: MHC expression in primary tumours versus metastases. Immunology. 2019;158(4):255–66.
32. Cornel AM, Mimpen IL, Nierkens S. MHC Class I Downregulation in cancer: underlying mechanisms and potential targets for cancer immunotherapy. Cancers (Basel). 2020;12(7).
33. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinf. 2016;54:1 30 1–1.
34. Uemura T, Shiozaki K, Yamaguchi K, Miyazaki S, Satomi S, Kato K, et al. Contribution of sialidase NEU1 to suppression of metastasis of human colon cancer cells through desialylation of integrin beta4. Oncogene. 2009;28(9):1218–29.
35. Forcella M, Mozzi A, Stefanini FM, Riva A, Epistolio S, Molinari F, et al. Deregulation of sialidases in human normal and tumor tissues. Cancer Biomark. 2018;21(3):591–601.
36. Jiao X, Liu W, Mahdessian H, Bryant P, Ringdahl J, Timofeeva M, et al. Recurrent, low-frequency coding variants contributing to colorectal cancer in the Swedish population. PLoS ONE. 2018;13(3): e0193547.
37. Bormann F, Stinzing S, Tierling S, Morkel M, Markelova MR, Walter J, et al. Epigenetic regulation of Amphiregulin and Epiregulin in colorectal cancer. Int J Cancer. 2019;144(3):569–81.
38. Cheng WL, Feng PH, Lee KY, Chen KY, Sun WL, Van Hiep N, et al. The role of EREG/EGFR pathway in tumor progression. Int J Mol Sci. 2021;22(23).
39. Song L, Peng X, Li Y, Xiao W, Jia J, Dong C, et al. The SEPT9 gene methylation assay is capable of detecting colorectal adenoma in opportunistic screening. Epigenomics. 2017;9(5):599–610.
40. Wang Y, Chen PM, Liu RB. Advance in plasma SEPT9 gene methylation assay for colorectal cancer early detection. World J Gastrointest Oncol. 2018;10(1):15–22.
41. Siegel EM, Berglund AE, Riggs BM, Eschrich SA, Putney RM, Ajidahun AO, et al. Expanding epigenomics to archived FFPE tissues: an evaluation of DNA repair methodologies. Cancer Epidemiol Biomark Prev. 2014;23(12):2622–31.
42. Dumenil TD, Wockner LF, Bettington M, McKeone DM, Klein K, Bowdler LM, et al. Genome-wide DNA methylation analysis of formalin-fixed paraffin embedded colorectal cancer tissue. Genes Chromosomes Cancer. 2014;53(7):537–48.
43. Espinal AC, Wang D, Yan L, Liu S, Tang L, Hu Q, et al. A methodological study of genome-wide DNA methylation analyses using matched archival formalin-fixed paraffin embedded and fresh frozen breast tumors. Oncotarget. 2017;8(9):14821–9.
44. Kling T, Wenger A, Beck S, Caren H. Validation of the MethylationEPIC BeadChip for fresh-frozen and formalin-fixed paraffin-embedded tumours. Clin Epigenet. 2017;9:33.

SUPPLEMENTARY MATERIAL

Supplementary Tables

Supp. Table 1 | Sample characteristics

	Normal (19)		Adenoma (17)		Carcinoma (19)	
Characteristic	*Number*	*%from total*	*Number*	*%from total*	*Number*	*%from total*
Sex						
Male	13	68,5	9	53	12	63
Female	6	31,5	8	47	7	37
Age						
Median	66,5	/	67,5	/	75,5	/
Interval	50 - 83	/	50 - 82	/	56-95	/
Location						
Left	11	57,9	10	59	11	57,9
Right	5	26,3	6	35	5	26,3
Unknown	3	15,8	1	6	3	15,8
Dysplasia						
Low-grade	/	/	15	88	/	/
High-grade	/	/	2	12	/	/
Type						
Tubulo-villous	/	/	5	29	/	/
Tubular	/	/	12	71	/	/

* This included paired samples from the same patient: 10 tumor-normal pairs and 1 adenoma-normal pair

Supp Table 2 | Samples that passed quality control of the in-house generated methylation data

Tissue type	Concentration (ng/µL)	Gender	Age
Nl	60,8	Male	83
LGA (Tb)	33,6	Female	70
CRC	187	Female	72
CRC	155	Male	71
HGA (Tv)	31,7	Male	64
Nl	22,7	Male	63
CRC	445	Male	76
LGA (Tv)	21,5	Female	50
LGA (Tv)	41,5	Male	70
CRC [1]	148	Female	62
LGA (Tb)	41,4	Female	82
LGA (Tb)	17,4	Female	68
CRC	118	Female	71
LGA (Tb)	23,2	Male	73
LGA (Tv)	114	Male	66
LGA (Tb)	31,8	Male	64
Nl [1]	63,2	Female	62
LGA (Tb)	20	Male	73
LGA (Tb)	23,4	Male	66
LGA (Tb)	46,3	Female	77
CRC	56,4	Female	72
LGA (Tb)	29,8	Male	67
CRC	29,1	Male	76
CRC [2]	298	Male	59
CRC	155	Female	68
Nl [2]	46,2	Male	59

(Nl = normal adjacent, CRC= colorectal cancer, LGA = low grade adenoma, HGA = high grade adenoma, Tb = tubular, Tv= tubulovillous, [1,2,...] = paired sample)

Supp. Table 3 | Overview of the datasets used for discovery, double evidencing and validation.

Dataset ID	Adenoma	Carcinoma	Normal	QC Passed Sum	Cohort
EMD_01	3	2	3	4	Validation
EMD_02	3	2	3	3	Validation
EMD_03	3	3	2	4	Validation
EMD_04	3	3	2	4	Validation
EMD_05	4	2	2	4	Validation
EMD_06	1	1	1	2	Validation
EMD_07	0	4	4	6	Validation
EMD_08	0	2	2	0	Validation
E-MTAB-6450	16	0	0	16	Discovery/DE
E-MTAB-7854	80	0	0	78	Discovery/DE
GSE132804	0	0	206	206	Discovery/DE
GSE151732	0	0	256	256	Discovery/DE
GSE166212	10	32	6	47	Discovery/DE
GSE199057	0	77	80	156	Discovery/DE
Sum	**123**	**128**	**567**	**786**	
GSE106556	0	0	10	10	DE
GSE129364	69	0	3	72	DE
GSE132804	0	0	128	128	DE
GSE139404	40	0	20	60	DE
GSE66555	0	0	43	34	DE
GSE68060	0	82	36	118	DE
GSE77718	0	96	96	191	DE
GSE77955	12	13	14	39	DE
TCGA_READ + COAD)	0	409	45	454	DE
Sum	**121**	**600**	**395**	**1106**	

EMD = experimental methylation data, E-MTAB= data from array express, GSE= data from GEO. Green = EPIC array data, blue = 450K array data.

Supp Table 4 | Location of DE DMPs in different groups within the genome

Group	Cg position	Chromosome	Gene	Feature-CGI	COSMIC-Census	Relevant references
A_C	cg02205193	13	/	IGR-opensea	/	/
	cg03211173	20	*MRGBP*	TSS200-island	/	/
	cg10059167	13	*OR7E156P*	TSS200-opensea	/	/
	cg11691825	1	/	IGR-opensea	/	/
	cg13822911	13	/	IGR-opensea	/	/
	cg14804531	11	/	IGR-shelf	/	/
	cg18363192	6	***NEU1***	TSS1500-shore	/	https://pubmed.ncbi.nlm.nih.gov/29278877/ AND https://pubmed.ncbi.nlm.nih.gov/29547645/
	cg19308222	4	*EREG*	TSS1500-opensea	/	https://pubmed.ncbi.nlm.nih.gov/34884633/
	cg20383155	6	***NEU1***	TSS1500-shore	/	https://pubmed.ncbi.nlm.nih.gov/29278877/ AND https://pubmed.ncbi.nlm.nih.gov/29547645/
	cg22630733	11	/	IGR-shelf	/	/
	cg22753340	6	**NEU1**	TSS1500-shore	/	https://pubmed.ncbi.nlm.nih.gov/29278877/ AND https://pubmed.ncbi.nlm.nih.gov/29547645/
	cg26487157	2	*PCBP1*	TSS1500-island	poly(rC) binding protein 1; CRC; missense mutations	/
	cg26642667	7	*SND1*	5'UTR-island	staphylococcal nuclease and tudor domain containing 1; pancreas acinar carcinoma; oncogene fusion; translocation partner = BRAF	/
A_N	cg00733780	10	*GAD2*	TSS200-island	/	https://pubmed.ncbi.nlm.nih.gov/22552777/
	cg00817367	12	*GRASP*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/23096130/
	cg01563031	11	*NELL1*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/35202405/
	cg01588438	8	*ADHFE1*	TSS200-island	/	https://pubmed.ncbi.nlm.nih.gov/35054365/ AND https://pubmed.ncbi.nlm.nih.gov/32317010/
	cg03807298	5	?	IGR-island	/	/
	cg04025964	1	*NR5A2*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/34253750/ AND https://pubmed.ncbi.nlm.nih.gov/30642095/
	cg04763554	1	*MIR137HG*	IGR-island	/	https://pubmed.ncbi.nlm.nih.gov/29730197/
	cg04921989	2	*GNAQP1*	IGR-island	/	/
	cg04996873	1	*KIAA1026*	Body-island	/	/
	cg06428620	5	*PCDHGA4*	Body-shore	/	/
	cg06454760	19	*ZNF135*	TSS200-island	/	https://pubmed.ncbi.nlm.nih.gov/35202405/
	cg06611358	19	*ZNF568*	IGR-island	/	https://pubmed.ncbi.nlm.nih.gov/28257124/
	cg06952671	2	*ITGA4*	5'UTR-island	/	https://pubmed.ncbi.nlm.nih.gov/35284127/
	cg07559273	14	*NOVA1*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/34305406/
	cg07690181	5	*EDIL3*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/21859567/

	cg07914084	1	*RYR2*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/35587572/
	cg07920503	13	*FAM123A*	1stExon-island	/	https://pubmed.ncbi.nlm.nih.gov/33396258/
	cg07923233	1	*SLC44A5*	IGR-island	/	/
	cg08452658	15	*TM6SF1*	TSS200-island	/	/
	cg08684639	4	*WDR17*	5'UTR-island	/	https://pubmed.ncbi.nlm.nih.gov/33580586/
	cg09100013	4	*UNC5C*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/34922605/
	cg09167553	1	*LRRC7*	IGR-island	/	https://pubmed.ncbi.nlm.nih.gov/33078631/
	cg09296001	7	*SND1*	Body-island	staphylococcal nuclease and tudor domain containing 1; pancreas acinar carcinoma; oncogenic fusion gene: translocation partner= BRAF	/
	cg10293925	7	*AMPH*	5'UTR-island	/	/
	cg10362542	17	*HRNBP3*	5'UTR-island	/	/
	cg10566121	11	*MPPED2*	5'UTR-island	/	https://pubmed.ncbi.nlm.nih.gov/30846004/
	cg11017065	11	*FLI1*	Body-island	Friend leukemia virus integration 1; Ewing sarcoma; oncogene, fusion; Translocation partner EWSR1	/
	cg11167100	15	*TM6SF1*	TSS200-island	/	/
	cg11767984	19	*ZNF788*	Body-island	/	/
	cg11855526	11	*MPPED2*	5'UTR-island	/	https://pubmed.ncbi.nlm.nih.gov/30846004/
	cg12821278	9	*DBC1*	TSS200-island	/	https://pubmed.ncbi.nlm.nih.gov/35801925/
	cg13267264	8	*PRDM14*	TSS200-island	/	https://pubmed.ncbi.nlm.nih.gov/35065650/
	cg13823136	1	*ST6GALNAC5*	TSS200-island	/	https://pubmed.ncbi.nlm.nih.gov/21400501/
	cg13895235	7	*PRKAR1B*	5'UTR-shore	/	https://pubmed.ncbi.nlm.nih.gov/8142263/
	cg14520424	1	*LRRC7*	IGR-island	/	https://pubmed.ncbi.nlm.nih.gov/33078631/
	cg14732324	5	*?*	IGR-island	/	/
	cg14817541	6	*EYA4*	5'UTR-island	/	https://pubmed.ncbi.nlm.nih.gov/32380793/
	cg15603568	11	*GRIA4*	5'UTR-island	/	https://pubmed.ncbi.nlm.nih.gov/34719006/
	cg15674193	2	*LRRFIP1*	TSS1500-island	/	https://pubmed.ncbi.nlm.nih.gov/22750095/
	cg15779837	19	*GRIN2D*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/33639954/
	cg16476975	7	*?*	IGR-island	/	/
	cg17301223	8	*OPLAH*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/35054365/
	cg17698295	8	*OPLAH*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/35054365/
	cg17872757	11	*FLI1*	Body-island	Friend leukemia virus integration 1; Ewing sarcoma; oncogene, fusion; Translocation partner EWSR1	/
	cg17892556	19	*ZNF625*	1stExon-island	/	https://pubmed.ncbi.nlm.nih.gov/25472652/
	cg18023283	12	*SLC6A15*	TSS1500-island	/	https://pubmed.ncbi.nlm.nih.gov/24485021/
	cg18412834	20	*FLJ16779*	TSS200-island	/	/

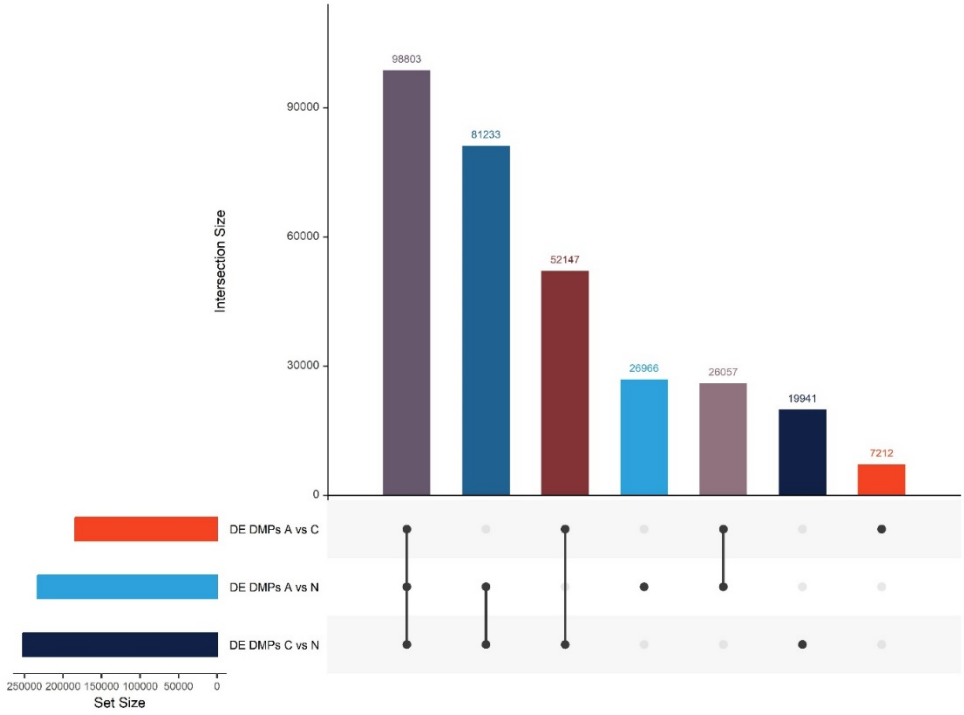
	cg18560328	11	*SYT9*	TSS200-island	/	/
	cg18601167	7	*PRKAR1B*	5'UTR-shore	/	https://pubmed.ncbi.nlm.nih.gov/8142263/
	cg18672939	10	*GFRA1*	1stExon-island	/	https://pubmed.ncbi.nlm.nih.gov/31988584/ AND https://pubmed.ncbi.nlm.nih.gov/33175846/
	cg21277995	6	*IRF4*	Body-island	interferon regulatory factor 4; MM; oncogene, TSG, fusion; Translocation partner of IGH	/
	cg21583226	2	*GALNT14*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/33387936/
	cg21995919	2	*ITGA4*	5'UTR-island	/	https://pubmed.ncbi.nlm.nih.gov/35284127/
	cg22834653	3	*FGF12*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/22552777/
	cg24171907	2	*CNRIP1*	1stExon-island	/	https://pubmed.ncbi.nlm.nih.gov/31833403/
	cg24446548	7	*TWIST1*	1stExon-island	/	https://pubmed.ncbi.nlm.nih.gov/35155261/
	cg24553673	1	*NR5A2*	Body-shore	/	https://pubmed.ncbi.nlm.nih.gov/34253750/ AND https://pubmed.ncbi.nlm.nih.gov/30642095/
	cg24857620	8	*NPBWR1*	TSS1500-island	/	/
	cg25570913	13	*?*	IGR-island	/	/
	cg26256223	8	*OPLAH*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/35054365/
	cg26560414	16	*CMTM3*	5'UTR-island	/	https://pubmed.ncbi.nlm.nih.gov/28782576/
	cg27200446	6	*MDFI*	5'UTR-island	/	https://pubmed.ncbi.nlm.nih.gov/35114976/
C_N	cg01612140	6	*enhancer*	IGR-opensea	/	/
	cg02272851	10	*enhancer*	IGR-opensea	/	/
	cg02420480	7	*?*	IGR-opensea	/	/
	cg03498081	16	*?*	IGR-opensea	/	/
	cg04456219	7	*DMR*	IGR-opensea	/	/
	cg04742334	20	*LOC339568*	Body-opensea	/	/
	cg04786142	1	*DMR*	IGR-opensea	/	/
	cg04810745	2	*enhancer*	IGR-opensea	/	/
	cg05433391	6	*DMR*	IGR-opensea	/	/
	cg05981038	5	*CARD6*	1stExon-opensea	/	https://pubmed.ncbi.nlm.nih.gov/20025480/
	cg06114334	7	*enhancer*	IGR-opensea	/	/
	cg06197966	10	*SVIL*	5'UTR-opensea	/	https://pubmed.ncbi.nlm.nih.gov/34213504/
	cg06433467	5	*DMR CTC537E7.3*	IGR-opensea	/	/
	cg06545761	16	*ZCCHC14*	3'UTR-island	/	/
	cg07474842	6	*MAP3K5*	Body-opensea	/	https://pubmed.ncbi.nlm.nih.gov/31910865/
	cg08550523	1	*enhancer*	IGR-opensea	/	/
	cg08584947	8	*enhancer*	IGR-opensea	/	/
	cg09129067	8	*?*	IGR-island	/	/
	cg09170112	12	*?*	IGR-opensea	/	/
	cg09287864	7	*DMR, enhancer*	IGR-opensea	/	/
	cg09841889	2	*enhancer*	IGR-opensea	/	/

	cg10077746	12	*DUSP6*	Body-shore	/	https://pubmed.ncbi.nlm.nih.gov/34790577/
	cg10344477	17	*B3GNTL1*	Body-shelf	/	/
	cg11838152	13	*ITGBL1*	Body-opensea	/	https://pubmed.ncbi.nlm.nih.gov/32211321/
	cg12060422	10	*KAZALD1*	1stExon-island	/	/
	cg12619262	7	*CHST12*	IGR-opensea	/	/
	cg13324103	10	*SVIL*	5'UTR-opensea	/	https://pubmed.ncbi.nlm.nih.gov/34213504/
	cg14094027	12	*PXN*	5'UTR-opensea	/	https://pubmed.ncbi.nlm.nih.gov/32705241/
	cg14642259	11	*MYBPC3*	Body-shore	/	/
	cg15536663	5	*EPB41L4A*	Body-opensea	/	https://pubmed.ncbi.nlm.nih.gov/30854042/
	cg16172099	8	*?*	IGR-shore	/	/
	cg16499677	14	*C14orf37*	Body-opensea	/	/
	cg16616514	4	*PPP2R2C*	3'UTR-opensea	/	https://pubmed.ncbi.nlm.nih.gov/34768523/
	cg17400812	7	*Enhancer*	IGR-opensea	/	/
	cg17494199	13	*?*	IGR-opensea	/	/
	cg18376288	1	*enhancer*	IGR-opensea	/	/
	cg19291696	13	*?*	IGR-opensea	/	/
	cg19945931	12	*?*	IGR-island	/	/
	cg20175702	15	*?*	IGR-opensea	/	/
	cg20873416	6	*HCG20*	IGR-opensea	/	/
	cg20979737	1	*?*	IGR-opensea	/	/
	cg21480725	15	*?*	IGR-opensea	/	/
	cg21536074	7	*GLI3*	Body-opensea	/	https://pubmed.ncbi.nlm.nih.gov/33506047/
	cg21769093	7	*VWC2*	Body-shelf	/	/
	cg22304399	19	*CACNA1A*	Body-island	/	https://pubmed.ncbi.nlm.nih.gov/35658861/
	cg22365276	15	*ANXA2*	5'UTR-shore	/	https://pubmed.ncbi.nlm.nih.gov/35163852/
	cg22743884	8	*DEFB135*	1stExon-opensea	/	/
	cg22963629	7	*?*	IGR-opensea	/	/
	cg24032190	15	*SMAD3*	Body-opensea	SMAD family member 3; colorectal carcinoma, oral squamous cell carcinoma; TSG; missense mutations	/
	cg25714865	7	*?*	IGR-shore	/	/
	cg25969107	15	*?*	IGR-opensea	/	/
	cg26450106	20	*CST7*	TSS1500-opensea	/	/
	cg26680608	6	*MAP3K5*	Body-opensea	/	https://pubmed.ncbi.nlm.nih.gov/31910865/
	cg26749518	19	*KLK9*	3'UTR-island	/	/
	cg27134730	7	*Enhancer*	IGR-opensea	/	/
	cg27365701	12	*DUSP6*	Body-shore	/	https://pubmed.ncbi.nlm.nih.gov/34790577/

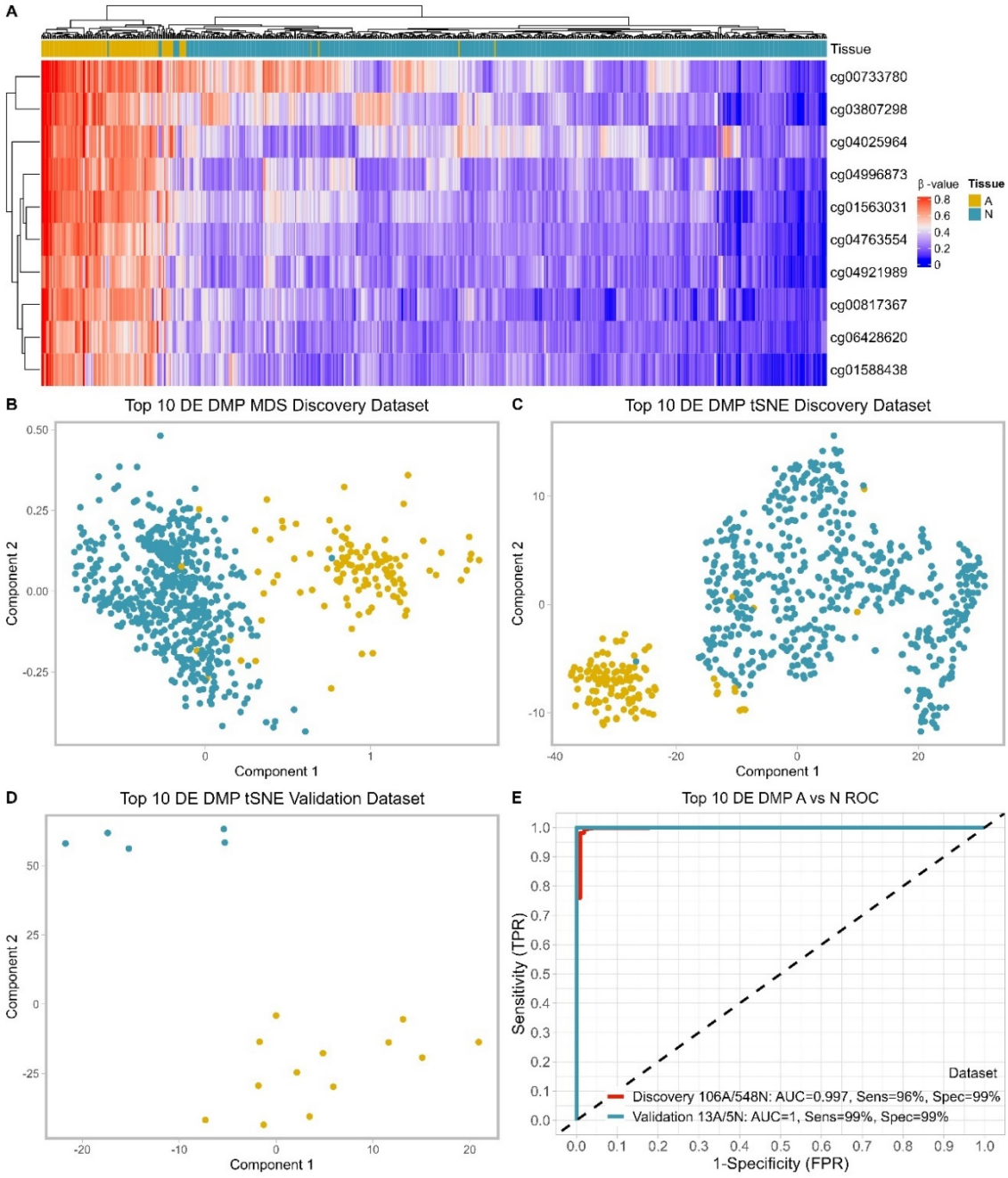
Suppl. Table 5 | Pathways in GSEA that contain the *NEU1* gene

Description	GSEA	ID
Lysosome	KEGG	4142
neutrophil activation involved in immune response	GO	GO:0002283
neutrophil mediated immunity	GO	GO:0002446
lysosomal membrane	GO	GO:0005765
vacuolar membrane	GO	GO:0005774
carbohydrate catabolic process	GO	GO:0016052
granulocyte activation	GO	GO:0036230
neutrophil activation	GO	GO:0042119
neutrophil degranulation	GO	GO:0043312
lytic vacuole membrane	GO	GO:0098852
carbohydrate derivative catabolic process	GO	GO:1901136
vacuolar lumen	GO	GO:0005775
specific granule	GO	GO:0042581
sphingolipid metabolic process	GO	GO:0006665
secretory granule lumen	GO	GO:0034774
Diseases of glycosylation	Reactome	R-HSA-3781865
Asparagine N-linked glycosylation	Reactome	R-HSA-446203
Neutrophil degranulation	Reactome	R-HSA-6798695

Supplemental figures

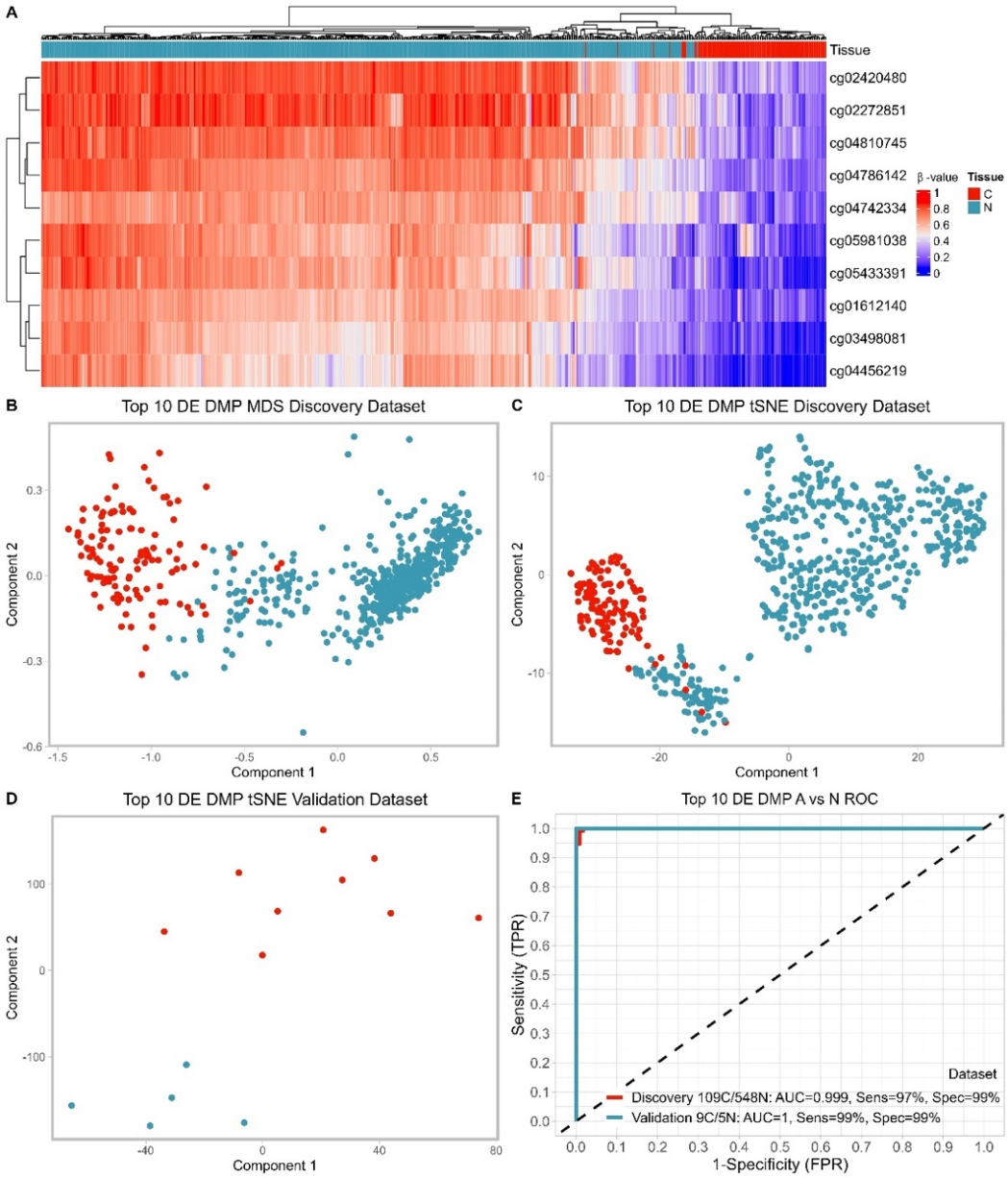


Suppl. Figure 1 | Counts of unique and common DE-DMPs in the three different colorectal tissue types. The upset diagram shows the total number of DE DMPs found in the discovery analysis prior to filtering based on |Δβ|. Legend: A= adenoma, N= normal, C= carcinoma



Suppl. Figure 2 | The top 10 selected DE DMP markers were effective at classifying adenomas and normal tissues. **A.** Heatmap and hierarchal clustering analysis of the discovery EPIC dataset based on the top 10 identified DE DMP markers shows a block like structure with most of the markers being hypermethylated in adenomas and hypomethylated in normal tissues. **B.** MDS clustering of the discovery dataset using the 10 markers shows 2 distinct clusters. **C.** tSNE clustering of the discovery dataset using the 10 markers could also resolve the two tissue types. **D.** tSNE clustering of the validation dataset using the 10 markers shows a clear separation between adenomas and normal tissues. **E.** ROC curves for the final top 10 DE DMP classifier model for both discovery and validation datasets from EPIC arrays. Sensitivity and specificity, for

distinguishing between adenomas and normal tissues, at various cut-off values for the datasets are plotted. The diagonal dotted line represents the line of no discrimination between tumor and normal tissues. DE DMP: double evidenced differentially methylated probes, ROC: receiver operating characteristic, MDS: multidimensional scaling, tSNE: t-distributed stochastic neighbor embedding, TPR: true positive rate, FPR: false positive rate, A: adenoma, N: normal tissue.



Suppl. Figure 3 |The top 10 selected DE DMP markers were effective at classifying carcinomas and normal tissues. **A.** Heatmap and hierarchal clustering analysis of the discovery EPIC dataset based on the top 10 identified DE DMP markers shows a block like structure with most of the markers being hypomethylated in carcinomas and hypermethylated in normal tissues. **B.** MDS clustering of the discovery dataset using the 10 markers shows 2 distinct clusters. **C.** tSNE clustering of the discovery dataset using the 10 markers could also resolve the two tissue types. **D.** tSNE clustering of the validation dataset using the 10 markers shows a clear separation between carcinomas and normal tissues. **E.** ROC curves for the final top 10 DE DMP classifier model for both discovery and validation datasets from EPIC arrays. Sensitivity and specificity, for distinguishing between carcinomas and normal tissues, at various cut-off values for the datasets are plotted. The diagonal dotted line represents the line of no discrimination between tumor and normal tissues. DE DMP: double evidenced differentially methylated probes, ROC: receiver operating characteristic, MDS: multidimensional scaling, tSNE: t-distributed stochastic neighbor embedding, TPR: true positive rate, FPR: false positive rate, C: carcinoma, N: normal tissue.

PART III | Creation of novel assays

CHAPTER 5

Simultaneous detection of eight cancer types using a multiplex droplet digital PCR assay

(Manuscript accepted in *Molecular Oncology* – DOI: 10.1002/1878-0261.13708)

Isabelle Neefs[1,2], Nele De Meulenaere[1,2], Thomas Vanpoucke[1,2], Janah Vandenhoeck[1,2], Dieter Peeters[3], Marc Peeters[2], Guy Van Camp[1,2] and Ken Op de Beeck[1,2]

1 Centre of Medical Genetics, University of Antwerp and Antwerp University Hospital, Prins Boudewijnlaan 43, 2650 Edegem, Belgium

2 Centre for Oncological Research Antwerp (CORE), University of Antwerp and Antwerp University Hospital, Universiteitsplein 1, 2610 Wilrijk, Belgium

3 Department of Pathology, Antwerp University Hospital, Drie Eikenstraat 655, 2650 Edegem, Belgium

ABSTRACT

Introduction: DNA methylation biomarkers have emerged as promising tools for cancer detection. Common methylation patterns across tumor types allow multi-cancer detection. ddPCR has gained considerable attention for methylation detection. However, multi-cancer detection using multiple targets in ddPCR has never been performed before. Therefore, we developed a multiplex ddPCR assay for multi-cancer detection.

Methods: Based on previous data analyses using TCGA, we selected differentially methylated targets for eight frequent tumor types. Three targets were validated using ddPCR in 103 tumor and 109 normal adjacent fresh frozen samples. Two distinct ddPCR assays were successfully developed. Output data from both assays is combined to obtain a read-out from the three targets together.

Results: Our overall ddPCR assay has a cross-validated AUC of 0.948. Performance between distinct cancer types varies, with sensitivities ranging from 53.8% to 100% and specificities ranging from 80% to 100%. Compared to previously published single-target parameters, we show that combining targets can drastically increase sensitivity and specificity, while lowering DNA input.

Conclusion: In conclusion, we are the first to report a multi-cancer methylation ddPCR assay, which allows for highly accurate tumor predictions.

INTRODUCTION

Cancer is one of the leading causes of death worldwide, with nearly 10 million deaths reported in 2020. The most common cancers are breast (2.26 million new cases), lung (2.21 million new cases), colorectal (1.93 million new cases) and prostate (1.41 million new cases) [1]. Many cancers can be cured when treated efficiently and, more importantly, detected early. However, patients are still frequently diagnosed in a late stage [2], despite efforts in screening programs and early detection methods. For example, 68% of lung cancer and 59% of colorectal cancer patients is diagnosed in stage IV [2], where 5-year survival rates are 7% and 14% respectively. There are currently only screening programs in place for colorectal, breast and cervical cancer in most western countries, and in some countries (e.g. the United States), screening for lung cancer is also recommended [3]. It is clear that there is much room to improve on cancer detection, especially in early stages. Therefore, early detection biomarkers have gained interest in recent years. DNA methylation is a very good candidate for such biomarkers. It occurs at the 5^{th} position of cytosines in the context of CpG dinucleotides. Tumor-associated modifications of the methylation status of CpG sites appear already early in carcinogenesis, possibly even before actual neoplastic transformation [4,5]. This makes DNA methylation changes the ideal target for early cancer detection. Methylation patterns are extensively altered between normal cells and cancer cells and are a very consistent feature as opposed to mutations, which typically vary at a wide range of sites [6,7]. Therefore, DNA methylation assays can be used off-the shelf, making them faster and cheaper to use compared to other assays. Distinct methylation patterns per tumor type have been observed and allow detection of tissue of origin. In addition, common methylation patterns exist across tumor types, which allow multi-cancer detection [7,8]. This multi-cancer detection has gained more attention in the past few years. Different assays are being developed and tested in clinical trials for safety and effectiveness [9,10]. The advantage of such multi-cancer detection tests is that they could facilitate the early detection of many cancers for which currently, no site-specific screening modality exists. Moreover, they could also detect cancers that are missed by the existing screening tests, as these assays are more sensitive than for example the existing imaging tests [10,11].

Just a decade ago, droplet digital PCR (ddPCR) was introduced as a highly sensitive method for detection and absolute quantification of targeted mutations in DNA [12,13]. The random partitioning of a clinical sample into thousands of droplets allows simultaneous PCR reactions in each individual droplet, drastically improving sensitivity [12,14,15]. In 2015, ddPCR was used for the first time to detect methylated alleles [16]. Since then, several studies have explored DNA methylation biomarkers in ddPCR assays for cancer detection and follow-up, mostly in liquid biopsies [6,7,13,17–20]. Up until this moment, there are no methylation ddPCR assays FDA-approved for use in clinical settings. None of the methylation ddPCR assays that are published, focused on multi-cancer detection using multiple (candidate) biomarkers. The multi-cancer tests that have been developed thus far, for example the Galleri® or PanSeer test usually make use of next-generation sequencing technologies

[21–23], although these technologies are inferior to ddPCR regarding detection sensitivity, more time-consuming and costly [13,18].

Previous research from our group has demonstrated the possibility of using DNA methylation biomarkers to discriminate eight different tumor tissues from each other and from normal tissue. The analyses were performed *in silico,* and our validated prediction model with 20 CpGs, common for the different cancer types, achieved 85% sensitivity at 91% specificity for stage one cancers [8]. However, efforts to design multiplex methylation assays often result in high-cost tests that are not affordable for routine diagnostics. Therefore, we aimed to develop and validate a triplex and duplex ddPCR assay for the detection of eight cancer types in fresh frozen tissue based on three differentially methylated targets.

METHODS

An overview of the methods is given in Figure 1.

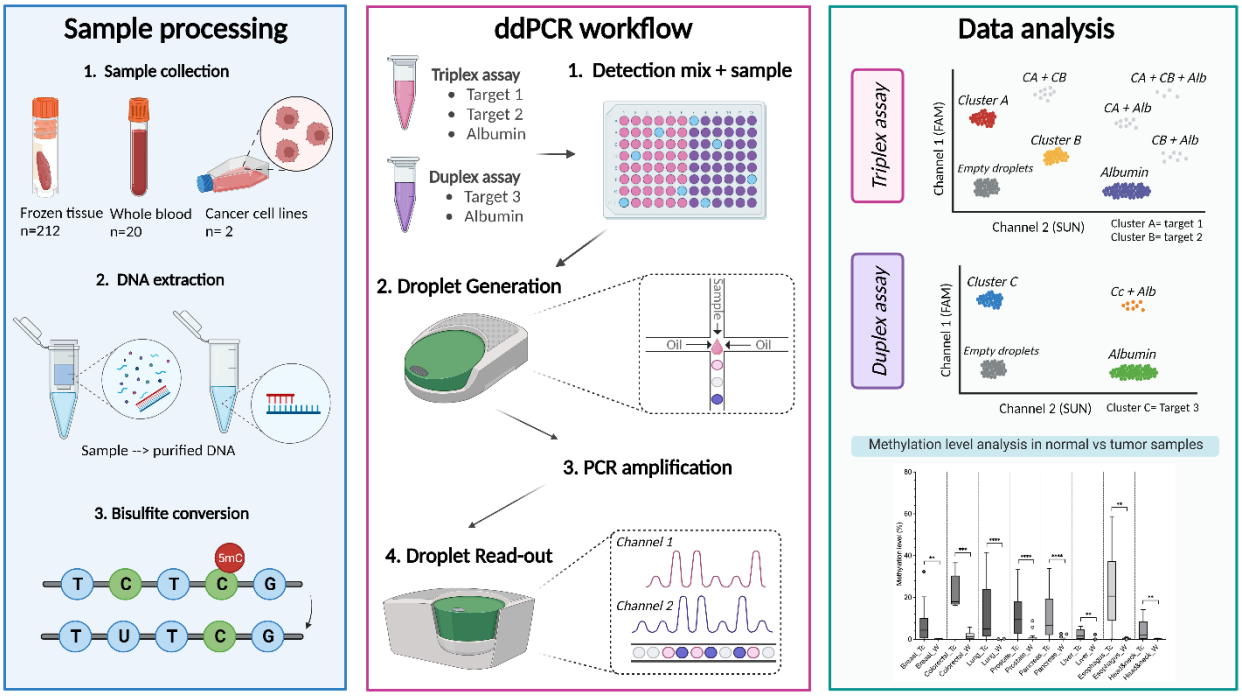


Figure 1 | Overview of the material and methods of the study**.** In the data analysis, clusters in grey indicate the double (or triple) positive clusters. Droplets of these clusters are excluded from the data

Sample collection and characteristics

Control samples

Whole blood samples (n= 20) were previously collected from healthy volunteers. They were used as negative controls and for determination of the limit of blank (LOB) and limit of detection (LOD) (see 2.5.1 Assay

characteristic measures for explanation). In addition, the HCT116 (CRC) and Cal27 (Head and neck) cancer cell lines were used as positive controls. They were cultured according to the standard protocols from the American Type Culture Collection (ATCC).

Fresh frozen tissue

Tumor tissue samples and normal adjacent tissue samples for eight cancer types were obtained from the Antwerp University Hospital biobank (UZA, Belgium). In total, 103 tumor and 109 normal adjacent tissue samples (Table 1) were collected and stored at -80°C. For all specimens, a hematoxylin and eosin-stained slide was made for microscopic analysis by the pathologist. Tissue type, presence of invasive tumor and overall tumor cell percentage (TcP) were verified. The average TcP was 40%, ranging from 5% to 95%. Samples from all invasive cancer stages (I-IV) were included in the study. The sample size calculation was performed as described in [24]. To determine the sample size, we used the target with the smallest effect size. A sample size with 7 cases and an equal number of controls holds 80% power to detect any difference between the tumor group (methylation level = 0.54 ± 0.23) and the normal group (methylation level = 0.18 ± 0.10), corrected for multiple testing (3 CpG sites) with a one-sided test. More details are given in supplementary table 2.

Table 1 | Tissue type and sample numbers used for ddPCR assays

Tissue type	Histological subtype	# Tumor	# Normal adjacent
Lung	Adenocarcinoma and squamous cell carcinoma	20	22
Breast	Invasive carcinoma	10	7
Colorectal	Adenocarcinoma	7	10
Pancreas	Adenocarcinoma	17	23
Liver	Hepatocellular carcinoma	12	10
Esophagus	Squamous cell carcinoma	10	5
Head and neck	Squamous cell carcinoma	13	8
Prostate	Adenocarcinoma	14	24
	Total	**103**	**109**

DNA extraction and bisulfite conversion

For whole blood samples, genomic DNA (gDNA) was extracted from 50ml blood using a standard salting-out process. The DNA was eluted in 1.5 ml TE buffer and stored at 4°C until further use. Genomic DNA from the cell lines was extracted from 10^6 cells using the Blood & Cell Culture DNA Maxi Kit (Qiagen, Hilden, Germany). DNA was eluted in 500 µL TE buffer and was stored at -20°C until further use. From fresh frozen tissue, genomic DNA was extracted from 1-10mg tissue using the QIAamp DNA Micro kit (Qiagen, Hilden, Germany). DNA was eluted in 70 µL TE buffer and stored at -20°C until further use. DNA concentrations were measured using the Qubit Fluorometer 2.0, DNA broad range assay (Invitrogen, MA, USA). Bisulfite conversion (20 ng per sample and per assay) was performed using the EZ DNA Methylation kit (Zymo Research) according to the manufacturer's instructions. The elution volume was set to 2µL x n + 1 µL, where n was the number of assays

(usually 2), and 1µl was added as a margin for technical errors. Bisulfite converted DNA was stored at -20°C and used within 10 days after conversion. We used 20 ng to have sufficient material for optimization and validation of the assays, in view of addressing potential issues e.g. rain in the assay.

Assay development

Target selection process

Targets were discovered through statistical analysis (pairwise comparison of cancer vs normal adjacent samples) of large datasets from The Cancer Genome Atlas. In total, 1 792 differentially methylated CpG sites were found. A total of 40 targets with a p-value of <0.01 and an absolute difference of 20% in tumor vs control methylation level were selected for primer design. This selection was based on 1) targets with the lowest p-value, 2) targets with the largest difference in methylation level and 3) targets for which primers could be designed. Of these 40 targets, the 27 best performing primer sets *in* silico were ordered from Integrated DNA technologies (IDT, Leuven, Belgium). Primers were evaluated using qPCR (384 CFX, Bio-Rad) on bisulfite converted gDNA from whole blood samples and cell lines. The targets with the largest differences in Cq values and the best matching melting temperatures were selected, which ultimately led to the retention of 3 targets for ddPCR assay development. Compared to the previous study [8], different targets were used due to technical reasons. For two of the probes that were previously described, primers could not be designed. For one of the probes, a SNP was present. For the last probe, the *in silico* primer conditions (melting temperature, %CG content, self-ligation...) were less optimal.

Primer and probe design for targets

Primers for target 1 were designed using *in silico* bisulfite conversion through the serial cloner software (version 2.6.1) and Primer3Plus (bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). Primers for targets 2 and 3 were designed using Meth Primer (urogene.org/cgibin/meth-primer/methprimer.cgi). Primers were ordered from Integrated DNA technologies (IDT, Leuven, Belgium). Probes were manually designed based on *in silico* bisulfite converted sequences. Non-fluorescent quencher, minor groove binder (NFQ-MGB) probes were used (IDT, Coralville, IA, USA). An overview of the newly designed primer and probe sequences can be found in Table 2. For albumin, primers were previously designed by Boeckx *et al.* [20]. The fluorophore from the published albumin probe was adapted from FAM to SUN, an equivalent for the VIC dye.

Table 2 | Newly developed primer and probe sequences for the triplex ddPCR assays

Target	1	2	3
Target name	*EMX1*	Chr5q14.1	*NXPH1*
Amplicon location	2:73147755-73147844+	5:76923876-76923965+	7:8482030-8482119+
Amplicon length	100 bp	87 bp	82 bp
Forward primer	5'CGAACGAAAAGGAATATGTTTG'3	5'GATACGTTTTTTTTGGAGAAGCGC'3	5'GAAGCGAAGGATTTTAGTTGTCG'3
Reverse primer	5'CTTCCAACGCCTCGATTAAC'3	5'CTTCATATCCCCAAACCCGAA'3	5'GAATACCCTCTCCTTCCGATATAACGA'3
Probe(s)	FAM-CGGCGCGGTTTCGGCG-N*	FAM-TGGGAGGTTTCGGGTATTTGAAGCG-N* SUN-TGGGAGGTTTCGGGTATTTGAAGCG-N*	FAM-CGTAGGGGGAGGTCGCGCG-N*

N*, NFQ-MGB = non-fluorescent quencher – minor groove binder

Optimization of the assays (temperature & concentration)

The assays were first assessed using a temperature gradient between 52°C and 62°C. The optimal temperature was chosen to further optimize the probe concentrations. Primer concentrations were adopted from Boeckx *et al.* [20]. In the triplex assay, target 1 was detected based on 6-FAM-fluorescence and target two was detected based on both 6-FAM and 6-SUN fluorescence. By combining the two different fluorophores on the same probe, a third color was artificially created for detection. By adjusting (increasing/decreasing) the ratio between 6-FAM and 6-SUN for the target that is detected in both channels, the clusters could be separated more (see also supplemental figure 3). In the duplex assay, target 3 was detected based on 6-FAM fluorescence. In both assays, albumin was used as a reference to determine the total amount of DNA present. Albumin detection was based on 6-SUN fluorescence, for which the probe concentration was adopted from Boeckx *et al.* [20]. After complete concentration optimization, a final temperature gradient around the previously chosen temperature was used to determine the optimal temperature for both assays.

ddPCR workflow

Assay composition

First, the assay mix containing primers, probes and Tris-HCl (pH 8) was made. Assay mixes were optimized as described above. Subsequently, the detection mix was composed of 11 µL 2x Bio-Rad ddPCR™ Supermix for probes (no dUTP), 1.1 µL of the assay mix, 2 µL DNA and 7.9 µL Milli-Q H_2O to a final volume of 22 µL. For both assays, the target and reference primer concentrations as well as the reference probe concentration were 900 nM in the final detection mix. Optimized concentrations for the target probes are given in the results section. Approximately 20,000 nanoliter sized droplets were generated according to the manufacturer's protocol using

the automated QX200™ Droplet generator (Bio-Rad). The resulting droplets were transferred to a 96-well plate and sealed. Target sequences were amplified using a Veriti™ thermal cycler (Applied Biosystems). Temperatures and times for the activation and inactivation steps were based on the recommended protocol from Bio-Rad for the ddPCR™ Supermix for Probes (no dUTP) (see Supplementary Table 1). Ramp rates of 2.5°C/sec were used based on Bio-Rad guidelines. The amplification temperature was set at the optimized temperature, given in the results section. An additional incubation step at 12°C for (at least) 15 minutes was added to the recommended Bio-Rad cycling protocol. Samples were immediately analyzed post amplification in the QX200™ Droplet Reader (Bio-Rad).

Sample analysis

After read-out, the data was analyzed using the QuantaSoft™ Analysis Pro software (version 1.0.596, Bio-Rad). This software was used to visualize the dispersion graphs. The positive control was used to set clusters based on fluorescence detection in the 6-FAM and/or 6-SUN channel. Four different clusters were manually assigned for the triplex assay (see Figure 2A). The different clusters are 1) empty droplets (shown in gray in Figure 2), 2) droplets containing the reference sequence albumin (shown in purple in Figure 2A), 3) droplets with the first target sequence (shown in red in Figure 2A) and 4) droplets containing the second target sequence (shown in yellow on Figure 2A). Similarly, three distinct clusters for the duplex assay were manually assigned (Figure 2B). Here, the reference sequence is given in green (Figure 2B) and target 3 is shown as blue droplets (Figure 2B). Droplets from double and triple positive clusters are excluded from all analyses, as the number of copies per target cannot be distinguished in these. A cut-off of at least 10,000 accepted droplets and at least 1,000 droplets for albumin (~4 ng as the lower limit of DNA input we accept) was set for analyzing all ddPCR experiments. The range of DNA varied from 4 ng to 19.72 ng, with an average of 9.47 ng. For samples in which the number of droplets was lower, the ddPCR protocol was repeated. Examples of a negative sample for both the triplex and the duplex assay is given in Supplementary Figure 1.

Assay characteristic measures

The limit of blank (LOB) is described as the number of false positive droplets that is detected in the negative controls (human genomic DNA) [25]. The LOB was used to discriminate the positive droplets (i.e. droplets containing the methylated target) from noise by subtraction of the LOB from detected positive droplets. The LOB was calculated as described in [20] and in Supplementary Table 2 . For both assays, the LOB was defined by the number of positive droplets measured in gDNA from whole blood samples that are not hypermethylated (n= 19 for the triplex assay, n=20 for the duplex assay). All data given in the results is normalized and corrected with the LOB (see also 2.5.1 normalization and methylation level). The limit of detection (LOD) is defined as 'the minimum concentration of the rare sequence that can be reliably differentiated from a negative control'

(see also Supplementary Table 2) [25]. To measure this minimal detectable methylation level, a DNA titration experiment using human methylated and non-methylated (WGA) DNA (Zymo Research, CA, USA) was performed. Five replicates of concentrations between 0% and 5% were measured with the assays to determine the LODs. They were calculated as described in [26] (see also supplementary Table 2). To determine the corresponding methylation level, linear regression was used. The minimally detectable levels, thus LODs, and are given in Table 3. The intra-assay coefficient of variability (%CV) demonstrates the variance between sample replicates within the assays and is a measure of repeatability. The copies per µL of four positive controls was used to calculate the %CV as described in Supplementary Table 2. The LOB, LOD and %CV of the targets is given in Table 3. The analytical detection sensitivity is a measure of the number of DNA copies that can be reliably detected. It was calculated by dividing the number of haploid genome equivalents (3) by the number of copies detected per well (in percentage, Supplementary Table 2) [25,27].

Normalization and methylation level

The number of positive and negative droplets was normalized to 20,000, the theoretical number of droplets that can be obtained using the QX200™ system. As shown in Supplementary Table 2, the number of accepted droplets was used for normalization.

The normalized number of droplets and the LOB were used to calculate the methylation percentage. This was calculated based on the number of droplets for the target sequences minus the LOB, and the reference sequence. After multiplying by 100, a percentage was obtained (see supplementary table 2). This represents the relative abundance of the methylated target sequence. The ideal marker would show a methylation level of 100% for the positive control and 0% for the negative control. The methylation level can vary according to the number of droplets for the target sequence or reference sequence that are assigned in the clusters. In exceptional cases, the methylation level can exceed 100%, for example when copy number variants are present in a sample.

Specificity and sensitivity

To calculate the specificity and sensitivity of the targets, first a receiver-operator characteristic (ROC) model was constructed and then validated using five-fold cross validation with the "ROCR" package in R (version 4.0.2). The cross-validated area under the ROC curve (cvAUC) was calculated to assess the predictive accuracy. Samples were considered methylated when the number of positive droplets for the target was higher than the LOB of that target. Samples were considered negative when (1) the number of positive droplets for the target was lower than the LOB and (2) no clear cluster for the target was observed. The sensitivities and specificities that were obtained per target and per cancer type are described in Table 4. For the combination of different targets, first the cutoff per target was determined based on the highest overall accuracy, dichotomizing the

results per target. Then, the separate predictions were combined by using the 'prediction' function of "ROCR", merged ROC curves were constructed and cvAUC values were calculated. When combining several targets, sensitivities and specificities can be calculated in different ways. For the triplex assay, a sample was considered positive if one target out of the two was methylated, which will further be described as the '1/2 threshold' (see also Fig. 4A). For the total ddPCR, which combines the three targets, two distinct thresholds were used. To achieve the highest sensitivity, a sample was considered positive based on methylation of one target out of three (further referred to as the '1/3 threshold'). For the highest specificity, a sample was only considered positive if the majority of the targets was methylated, i.e. two out of three (referred to as the '2/3 threshold').

Statistical analysis

The GraphPad prism (version 9.5.1) software was used to perform statistical tests along with the generation of figures. To determine whether the difference that was observed between the methylation levels of DNA from normal adjacent samples and tumor samples was significant, the Mann-Whitney U test was used. To analyze whether there are differences in methylation levels between cancer stages, the Kruskal-Wallis test was used. Post-hoc analyses were performed using the Mann-Whitney U test.

RESULTS

Target selection and assay development

Previous bioinformatic analyses from TCGA by our group [8] demonstrated the use of differentially methylated CpG sites for discrimination between controls and different tumor types. From this study, several CpG targets were further investigated for their application as biomarkers. A total of three targets was retained. Primers and probes for the selected targets were first analyzed using qPCR (CFX384, Bio-Rad) to determine their ability to discriminate hypermethylated samples (cancer cell lines) from hypomethylated samples (healthy blood samples). Two cell lines (HCT116 and Cal27) and two blood samples were analyzed in duplicate. The average Cq was used to determine the ΔCq. The qPCR results of these targets are displayed in Supplementary Table 3. The minimal ΔCq was 4.6, while the maximal ΔCq was 11.3. Based on the three targets, two ddPCR assays were developed. ddPCR was the chosen method, because its detection sensitivity is crucial for the future application of the assays in liquid biopsies. Also, around 13% of samples have a low TcP (<10%), for which we believe higher analytical sensitivity is beneficial. The ddPCR assays were optimized through temperature gradients and concentration gradients of probes. The first assay combines targets 1 and 2. For this assay, the optimal temperature was 55°C. The optimal concentrations for probes were 450 nM for target 1, 680 nM for target 2 – FAM and 1,4 µM for target 2 – SUN in the final detection mix. The positive control (HCT116) had an LOB-corrected methylation percentage of 64.30% for target 1 and 71.62% for target 2 (Figure 2A). The second assay

consists of target 3. For this assay, the optimal temperature was 58°C. The optimal concentration for the probe was 2.93 µM in the final detection mix. HCT116 has an LOB corrected methylation percentage of 73.79% (Figure 2B).

Figure 2 | Dispersion graphs of the triplex and duplex assay. The graphs show the dispersion of droplets for a positive control (HCT116, CRC cell line). A) Triplex assay consisting of target 1, 2 and reference Albumin. B) Duplex assay consisting of target 3 and reference albumin.

Characteristics of the assays

For the different targets, several measures (LOB, LOD and others) were calculated as described in section 2.5.1 and Supplementary Table 2. An overview of the parameters is given in Table 3. The %CV indicates great repeatability for targets 1 and 2 and good repeatability for target 3. In both assays, the average analytical detection sensitivity for the reference target was calculated over all samples (Supplementary Table 2). For the triplex assay, the average detection sensitivity was 0.11% ± 0.04%. For the duplex assay, a very similar average detection sensitivity of 0.10% ± 0.05% was found, indicating an equally good performance for both assays. Finally, the number of accepted droplets and copies per droplet measured for the assays was evaluated. In the triplex assay, an average of 14,251 ± 1,932 accepted droplets were found, with an average of 147 ± 86 albumin

copies per droplet (Poisson-corrected minimum) per run. For the duplex assay, the average of accepted droplets was 13,315 ± 1,772 with on average 180 ± 121 albumin copies per droplets (Poisson-corrected minimum) per run.

Table 3 | Limit of blank, Limit of detection and inter-assay coefficient of variability for the targets

Target	LOB (droplets)	LOB (meth %)	LOD (droplets)	LOD (meth %)	%CV
1. *EMX1*	2.8	0.31	7.2	0.84	4.3
2. Chr5q14.1	10.5	0.84	17.3	1.45	2.2
3. *NXPH1*	6.0	0.94	11.6	1.75	8.6

Methylation analysis of the targets in fresh frozen tumor and normal adjacent tissue

The methylation levels of 103 tumor and 109 normal adjacent fresh frozen tissues were analyzed using both the triplex and the duplex assay. Methylation levels are summarized in Figure 3. For target 1, all tissue types except colorectal tissue (p= 0.541) show a significant difference in methylation percentages between normal and tumor tissue. Lung, prostate and pancreatic tissue give the most significant results with p-values below 0.0001. For target 2, all tissues have a significant difference in methylation levels. As for target 1, lung, prostate and pancreatic cancer give the most significant results again. For target 3, all tissues give significant differences in methylation levels for tumor samples compared to normal adjacent samples as well. Here, lung and pancreatic cancer have the most significant results, followed by colorectal and esophageal cancer.

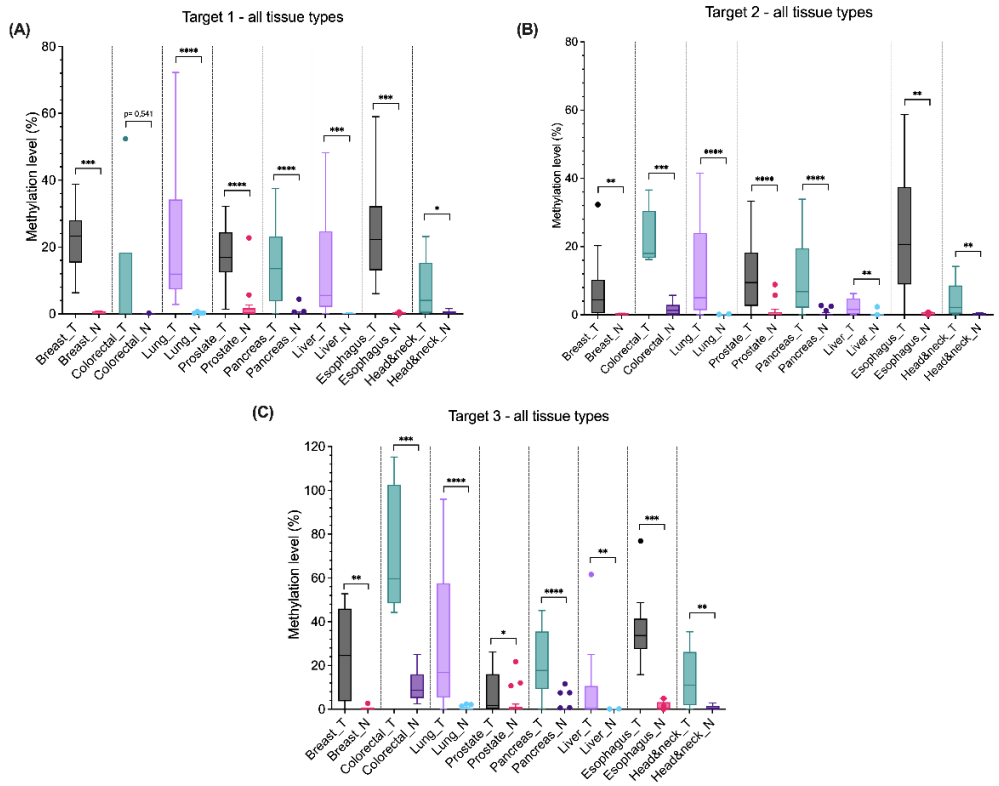


Figure 3 | Overview of methylation levels in tumor and normal adjacent fresh frozen samples. A) Overview for target 1. B) Overview for target 2. C) Overview for target 3. P-values are indicated with asterisks, where *= $p < 0.05$, **= $p < 0.01$, ***= $p < 0.001$ and ****= $p < 0.0001$. Tc= Tumor, N= normal adjacent. Mann-Whitney U tests were performed using GraphPad Prism.

Methylation levels in tumor tissue show a wide range from 0% to 72.2% for target 1, 0% to 58.7% for target 2 and 0% - 115% for target 3. In contrast, the normal adjacent tissue's range is limited from 0% to 5.7% for both targets 1 and 2 and 0% to 16.7% for target 3, excluding one outlier in the prostate group (22.7% for target 1, 8.9% for target 2 and 21.8% for target 3) and one in the colorectal group for target 3 (25%).

There is no strong correlation of the methylation levels between the different triplex targets for the same sample (r= 0.59 for tumor samples and r= 0.71 for normal samples, Supplementary Figure 2A-B). Furthermore, there is no correlation between the tumor cell percentage and the methylation levels of the tumor samples (r= 0.22 for target 1, r= 0.14 for target 2 and r= 0.12 for target 3, Supplementary Figure 2C-E). Last, there is no significant difference in methylation levels between different cancer stages for target 1 and target 3. For target 2, there is only a significant difference found in methylation levels between stages 1 and 3 (p= 0.03) (Supplementary Figure 2F-H). The diagnostic performance of the targets per cancer stage was evaluated by calculating the sensitivity from ROC analyses and is given in Supplementary Table 5.

Sensitivity and specificity of the targets for multi cancer detection

To evaluate sensitivities and specificities for the targets, cross-validated ROC plots were generated per tissue type, per target and for the combination of targets. Data from the separate targets is shown in Table 4 and ROC plots for the combination of the targets are shown in Figure 4. They were used to determine the best threshold for sensitivity and specificity. For the combination of targets, samples were analyzed based on a 1/2, 1/3 and 2/3 threshold, as explained in the methods. A summary of the parameters can be found in Table 4.

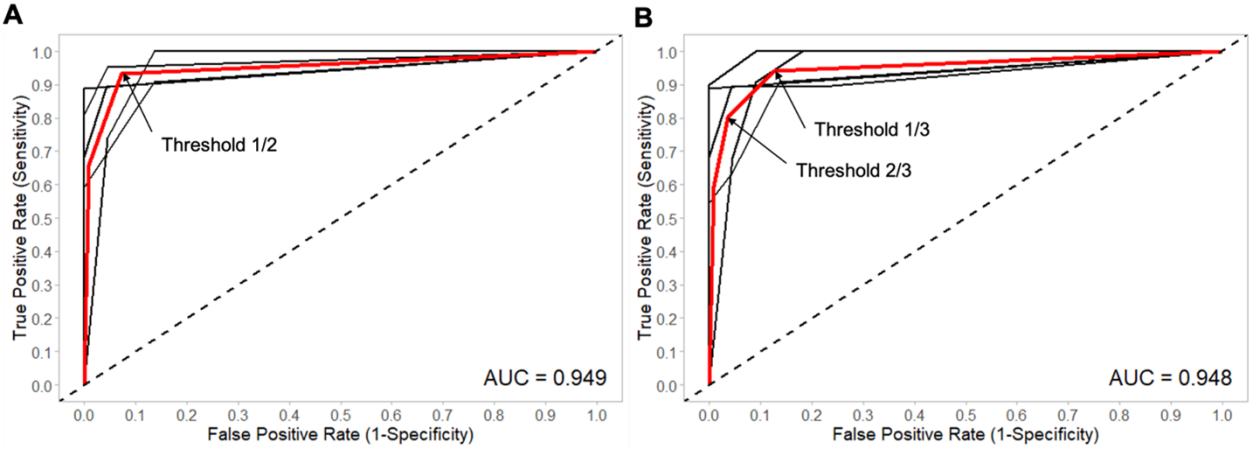


Figure 4 | Receiver Operator Characteristic (ROC) curves. A) ROC plot for the two targets of the triplex assay. B) ROC plot combining all targets. AUC = (cross-validated) area under the curve. For each target, the 5-fold cross validation ROC curves (black) and the mean ROC curve (red) is plotted. Breakpoints reflect the number of targets agreeing in the classification. Depending on the numbers of targets (i.e. breakpoints) that classify tumor and normal samples correctly, the sensitivity and specificity varies. Arrows indicate the breakpoints (i.e. thresholds) that were used for calculations of the sensitivities and specificities.

In the multi-cancer context, numbers indicate that, although targets perform well separately (see Table 4), combining targets can increase the sensitivity and specificity. For example, the overall sensitivity of target one is 82.5%, and for target 2, it is 76.6%. However, a sensitivity of 93.2% is reached when looking at both targets together. The cvAUC for the triplex is 0.949. The same is true for the overall sensitivity of target three, which is 74.3% but in the combination of all three targets, the overall sensitivity reaches 94.1% (1/3 threshold). The changes in specificity are less drastic, but also underline the strength of combining targets to improve the diagnostic value (see Table 4). Adding the third candidate-biomarker target to the model, the total cvAUC reaches 0.948. Depending on the use of the 1/3 or 2/3 threshold, the model performs better for sensitivity or specificity respectively (see Table 4). Compared to the previously described *in silico* analyses by Ibrahim *et al.* [8], all three targets, the triplex and overall assay perform similarly (supplementary table 4).

Looking at the cancer types separately, there are several cancer types for which the targets reach 100% sensitivity and 100% specificity. In contrast, some targets do not perform well in specific cancer types (see Table 4). However, as is also described above, combining at least two targets increases the sensitivity and specificity drastically (see Table 4).

Table 4 | Cross-validated AUC and sensitivity-specificity based on ROC analyses

		Overall	**Breast**	**Colorectal**	**Lung**	**Prostate**	**Pancreas**	**Liver**	**Oesophagus**	**Head and neck**
Target 1	*Sens*	**82.5%**	100%	28.6%	100 %	100%	82.4%	66.7%	100%	53.8%
	Spec	**98.2%**	100%	100%	100 %	91.7%	100%	100%	100%	100%
Target 2	*Sens*	**76.6%**	60%	100%	85%	78.6%	76.5%	66.7%	80%	69.2%
	Spec	**93.6%**	100%	70%	100 %	91.7%	91.3%	100%	100%	100%
Triplex	*Sens*	**93.2%**	100%	100%	100 %	100%	88.2%	83.3%	100%	76.9%
	Spec	**92.7%**	100%	70%	100 %	87.5%	91.3%	100%	100%	100%
	cvAUC	**0.949**	1	0.893	1	0.966	0.928	0.917	1	0.885
Target 3	*Sens*	**74.3%**	80%	100%	85%	57.1%	81.3%	41.7%	100%	61.5%
	Spec	**91%**	100%	40%	100 %	91.7%	91.3%	100%	100%	100%
Total, 1/3	*Sens*	**94.1%**	100%	100%	100 %	100%	87.5%	83.3%	100%	84.6%
	Spec	**87.3%**	100%	30%	100 %	83.3%	87.0%	100%	100%	100%
Total, 2/3	*Sens*	**80.2%**	80%	100%	90%	85.7%	81.3%	58.3%	100%	53.8%
	Spec	**96.4%**	100%	80%	100 %	95.8%	95.7%	100%	100%	100%
Total	*cvAUC*	**0.948**	1	0.93	1	0.96	0.921	0.917	1	0.923

1/3 = sample is classified as tumor if there is a positive signal for one of the three targets, 2/3 = sample is classified as tumor if there is a positive signal for the majority of the targets (two out of three), sens= sensitivity, spec= specificity, cvAUC= cross-validated Area Under the Curve.

DISCUSSION

In this study, we successfully established two multi-cancer detection ddPCR assays. Previous bioinformatic analyses by our group [8] have already demonstrated the use of differentially methylated CpG sites for discrimination between controls and multiple tumor types. From the analyses, multiple differentially methylated CpG sites were investigated. Selecting the sites based on the p-value and discriminatory power (at least 20% difference in methylation level between tumor and normal samples) was an unbiased approach that left us with several candidate-biomarkers. As ddPCR was the preferred method in view of future applications in liquid biopsies, several attempts in primer and probe design and testing have led to the ultimate retention of three targets.

Although the biomarkers used in our assays were selected purely on differential methylation, we investigated in which genes they are located. The first target covers a total of 17 differentially methylated CpGs in its amplified product. These sites are in the *EMX1* (Empty Spiracles-Like Protein 1) gene [28]. Although this gene has no clear role in any carcinogenic processes, it has been reported to be differentially methylated in both gastric and hepatocellular carcinomas (HCC) [29–32]. The second target covers 6 differentially methylated CpGs in total, in an intergenic region located on the short arm of chromosome 5 (Chr5q14.1). Loss of the chr5q arm has been described in myelodysplastic syndrome and several cancers [33–36], but methylation at this specific location has not yet been described. In the third target region, a total of 9 differentially methylated CpGs were found. The gene at this location is the Neurexophilin 1 (*NXPH1*) gene. It is known to be involved in acute myeloid lymphoma [28]. Furthermore, it was described as a potential methylation biomarker in breast cancer [37] and intraductal papillary mucinous neoplasms (IPMNs) of the pancreas [38].

For the ddPCR analyses, two different types of assays were developed. In the triplex assay, target 1 has an LOB of 3, which is comparable with published ddPCR assays [39]. The %CV is low (4.3), indicating limited variability within the assay. Target 2 has a higher LOB and lower %CV (11 and 2.2 respectively). Target 2 and 3 have a rather high LOD (methylation level of 1.45% – 1.75%), meaning that sufficient methylated DNA is needed for reliable detection of this target. Although all measures are within acceptable limits, the rather high LOB and LOD might hamper the application potential of this target in samples with very low tumor fractions (e.g. early stage liquid biopsies). Target 3 has an intermediate LOB of 6 and a %CV of 8.6, which again indicates good agreement within the assay. All targets have a standard error of the mean of approximately 1.5, indicating good repeatability. Last, the analytical detection sensitivity of the triplex assay and duplex assay is very similar (0.10% and 0.11%), indicating that both assays perform equally well. In the analyses, clusters were assigned manually for both assays. This remains standard practice for multiplex assays and the characterization parameters did not implicate this is as a problem. The manual assignment of the clusters can slightly affect the absolute quantification. However, for this assays' purpose, the relative quantification (i.e. methylation level) is

of importance. Using manual assignment of clusters to have positive droplets only in the designated cluster, allowed us to minimize the LOB of the targets and as such minimize the number of false positive droplets. This is more important for the clinical implementation than the absolute quantification, in this specific assay context. Nevertheless, manual assignment could make standardization among different laboratories more difficult, and standardized methylation analyses are necessary for implementation in the clinic. In this view, Jeanmougin *et al.* described PoDcall, an R package for automated calling of positive droplets, quantification and normalization of methylation levels [40]. However, the paper was only published in 2022, so it might need more testing and validation before it can be implemented in other studies as a standard tool. Furthermore, in these analyses, we only verify positive clusters. This might lead to a minor decrease in sensitivity, which might be unfavorable for cfDNA analysis. However, Whale et al. previously described that unbiased estimates can be achieved by using either all partitions or a subset of partitions, as long as no linkage between the targets occurs [41]. Since our targets are located on different chromosomes, they are not linked and we can expect unbiased results by using only positive clusters.

The optimized triplex and duplex ddPCR assays were used to assess fresh frozen tissue samples. For all tissue types and targets, except target 1 in colorectal tissue, a significant difference between tumor and normal adjacent samples was found. Lung tissue and pancreatic tissue had the most significant results in all targets. As the tumor cell percentages and cancer stages were in the same range for all tissue types, this is most likely due to their larger sample size (n= ~20) compared to other tissue types (n= ~10). In most of the normal tissue groups, the methylation levels are not 0% but vary. Normal adjacent tissue was used, which was taken at a certain (unknown) distance from the tumor. Field cancerization could be a possible explanation for higher methylation levels in normal adjacent tissue [42]. This is most likely for target 3 in colorectal normal tissue, where methylation levels vary from 2.5% to 25%. Nonetheless, for this target, the colorectal tumor samples also show high methylation levels, making it possible to perfectly distinguish between tumor and normal samples. Head and neck tissue remains the most difficult tissue type with the lowest sensitivity (see Table 4) compared to other tissue types. DNA methylation is tissue type specific, and head and neck cancers comprise several tissues (mouth, sinuses, nose and throat). Importantly, the etiology of the different sub-groups in head and neck cancer varies. These reasons could possibly explain why the targets are not as sensitive in head and neck cancers as in other tissue types. Finally, for esophagus cancer the number of normal adjacent tissues (n=5) does not meet the criteria from the sample size calculation (n= at least 7). Further analyses with more samples are warranted to draw robust conclusions for this subgroup.

Although we expected to find a correlation between the estimated TcP and the methylation level per target, there was none. As outlined in the material and methods section, HE-slides were reviewed by a pathologist (D.P.) to verify the tissue type and estimate the percentage of tumor cells. Since this material is precious, only one slide was used for verification. Unfortunately, estimating the tumor cell percentage stays a bit arbitrary

and subjective. Moreover, the presence of other cells within the tumor samples, such as normal cells and immune cells (originating from inflammatory processes), dilute the number of tumor cells which affects the estimated TcP to an uncertain extent. Lastly, this correlation was made between the methylation level of one location in the epigenome and the overall TcP of the tumor sample, which might hamper the potential to find a correlation.

In literature, methylation specific ddPCR assays typically involve the detection of one target gene in one cancer type. Almost all papers describe the use of ddPCR assays in both tissue and liquid biopsies (cfDNA) [6, 7,14,17]. In more recent research, the combination of several targets in separate ddPCR duplex assays is often described [17]. However, by creating separate assays, more DNA input is needed compared to our triplex assay. Our triplex assay could be beneficial for low DNA input samples or samples with limited tumor fractions such as cfDNA, as two targets can be investigated using the same amount of DNA. Compared to qPCR, ddPCR has a better analytical detection sensitivity (1% vs 0.001%) which is important for cfDNA or FFPE samples and early detection. As observed in our experiments (see supplementary text), the LODs of ddPCR were far better than those of the qPCR (~1% versus ~6% respectively). In view of future clinical applications, which can be both FFPE and liquid biopsy related, we designed ddPCR assays and optimized them using fresh frozen material. The advantage is that both ddPCR assays are now readily available and validated.

Furthermore, combining targets in one assay increases the sensitivity as can be appreciated by comparing data from Table 4 (targets 1 and 2 versus triplex). Table 4 shows that not all targets perform equally well in all cancer types. However, we use this as a strength of our overall assay. By combining all three targets from two ddPCR assays, we were able to achieve high sensitivities and specificities for all cancer types. To allow for different applications, both the 1/3 and 2/3 thresholds were used to calculate diagnostic parameters. As seen in Table 4, the sensitivity is higher when using the 1/3 threshold approach. In contrast, the 2/3 threshold approach allows for higher specificity.

The use of multiplex ddPCR assays in cancer has been described for microsatellite instability, mutation and copy number alteration detection [43 - 48]. Methylation detection has been used more in other fields such as detection of fetal fraction in blood samples and detection of SARS-CoV-2 [49,50]. There is one cancer multiplex methylation ddPCR assay described in literature, which was used to detect differences in methylation before and after neoadjuvant therapy in breast cancer [51]. Here, they used different concentrations of the FAM probe for the detection of two target genes in one assay. In addition, there is only one more assay described very recently by Zhao *et al*. for lung cancer screening [52]. Despite their capability of combining four targets into one assay, the use of their own microfluidics device limits broad applications in the clinic.

Regarding multi-cancer applications, we are the first to describe a multi-cancer ddPCR assay using multiplexing. There is one recent paper by Manoochehri *et al*. that describes the use of the *SST* gene as an

interesting marker for several cancer types, including all types described in our study [53]. However, only pancreatic cancer tissue was evaluated *in vitro* with a ddPCR assay. The AUC for this was 1. The combination of our three targets reached an AUC of 0.957 for pancreatic tissue and is thus comparable to the *SST* biomarker performance. The other cancer types have only been described in an *in-silico* test for which no AUCs were calculated, so we cannot compare the *SST* gene to our targets for multi-cancer detection [53]. Other multi-cancer detection assays described in literature, e.g. the Galleri test from GRAIL, are NGS based [21]. However, in view of implementation in the clinic, ddPCR is less labor intensive and more cost-efficient than NGS [18]. Multi-cancer assays often also incorporate the detection of tissue of origin (TOO), which our assay does not specify.

In conclusion, we are the first to report a multi-cancer multiplex methylation ddPCR assay. The overall assay with three methylation biomarkers reached an AUC of 0.948 in 8 different tumor types. In the future, more multiplexing is likely to be achieved with the novel QX600 system of Bio-Rad, where 6 fluorescent channels will allow multiplexing of approximately 10 targets, which could greatly enhance sensitivity. Furthermore, the assays need to be evaluated in FFPE material and liquid biopsies to assess their performance in (poor quality) DNA from FFPE samples and cfDNA, as this will be of more interest to be implemented in the clinic. The results of this study can serve as a solid basis for further MCED-ddPCR cfDNA applications.

DECLARATIONS

Acknowledgements

We would like to thank all patients and the UZA biobank for their contribution to this study. Furthermore, we want to express our gratitude to our lab technician Anne Schepers for her work related to the processing of the fresh frozen samples.

Author's contributions

The work reported in the paper has been performed by the authors, unless clearly specified in the text **IN**: Conceptualization of the study, methodology, interpreting results, writing original draft, reviewing and editing of the final draft. **NDM**: Methodology, interpreting results, reviewing and editing of the final draft. **JV**: Conceptualization of the study, interpreting results, reviewing and editing of the final draft. **TVP**: Conceptualization of the study, interpreting results, reviewing and editing of the final draft. **DP**: Methodology (histopathological analysis), reviewing and editing of the final draft. **MP**: Conceptualization of the study, reviewing and editing the final draft. **GVC**: Conceptualization of the study, interpreting results, reviewing and

editing of the final draft. **KODB**: Conceptualization of the study, interpreting results, reviewing and editing the final draft. All authors have read and agreed to the published version of the manuscript.

Ethics approval and consent to participate

This study was conducted in accordance with Good Clinical Practice guidelines and the Declaration of Helsinki. Tissue samples were retrospectively collected from the Biobank of the Antwerp University Hospital for use in this study. Patients have given consent for the use of their bodily material in research when they consented to the invasive procedure (according to article 20 of the Belgian Law on the procurement and use of human corporal material intended for human application or scientific research of 19 December 2008). All healthy volunteers have given written consent for the use of their bodily fluid in research. The UZA ethical committee approved the use of all samples (Ref. N°20/02/056 and Ref. N°41/14/426).

Data availability

All data generated or analyzed during the study are included in this published article and its Supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding information

Research in the lab of Prof. Guy Van Camp is supported by grants awarded by the University of Antwerp (IOF/SBO 43782, BOF/Methusalem grant 40790, BOF/TOP 39705). Isabelle Neefs is supported by a strategic basic PhD fellowship of the Research Foundation Flanders (FWO; IS67523N).

REFERENCES

1. WHO. Cancer - Screening and early detection. 2010.

2. National Cancer Institute. Stage at Diagnosis. 2022.

3. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023;73:17–48.

4. Kulis M, Esteller M. Chapter 2 - DNA Methylation and Cancer. Epigenetics and Cancer, Part A. 2010;70:27–56.

5. Constâncio V, Nunes SP, Moreira-Barbosa C, Freitas R, Oliveira J, Pousa I, et al. Early detection of the major male cancer types in blood-based liquid biopsies using a DNA methylation panel. Clin Epigenetics. 2019;11.

6. Cho N-Y, Park J-W, Wen X, Shin Y-J, Kang J-K, Song S-H, et al. Blood-Based Detection of Colorectal Cancer Using Cancer-Specific DNA Methylation Markers. 2020; Available from: https://doi.org/10.3390/diagnostics

7. Wang D, O'Rourke D, Sanchez-Garcia JF, Cai T, Scheuenpflug J, Feng Z. Development of a liquid biopsy based purely quantitative digital droplet PCR assay for detection of MLH1 promoter methylation in colorectal cancer patients. BMC Cancer. 2021;21.

8. Ibrahim J, Op de Beeck K, Fransen E, Peeters M, Van Camp G. Genome-wide DNA methylation profiling and identification of potential pan-cancer and tumor-specific biomarkers. Mol Oncol. 2022;16:2432–47.

9. National Cancer Institute (NCI). Multi-Cancer Detection (MCD) Research. https://prevention.cancer.gov/major-programs/multi-cancer-detection-mcd-research.

10. Farooq M, Leevan E, Ahmed J, Ko B, Shin S, De Souza A, et al. Blood-based multi-cancer detection: A state-of-the-art update. 2024 [cited 2024 Feb 1]; Available from: https://doi.org/10.1016/j.currproblcancer.2023.101059

11. Myers RE, Hallman MH, Shimada A, DiCarlo M, Davis K, Leach WT, et al. Primary Care Patient Interest in Multi-Cancer Early Detection for Cancer Screening. J Pers Med. 2023;13.

12. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal Chem. 2011;83:8604–10.

13. Holm M, Andersson E, Osterlund E, Ovissi A, Soveri LM, Anttonen AK, et al. Detection of KRAS mutations in liquid biopsies from metastatic colorectal cancer patients using droplet digital PCR, Idylla, and next generation sequencing. PLoS One. 2020;15.

14. Pharo HD, Andresen K, Berg KCG, Lothe RA, Jeanmougin M, Lind GE. A robust internal control for high-precision DNA methylation analyses by droplet digital PCR. Clin Epigenetics. 2018;10.

15. van Wesenbeeck L, Janssens L, Meeuws H, Lagatie O, Stuyver L. Droplet digital PCR is an accurate method to assess methylation status on FFPE samples. Epigenetics. 2018;13:207–13.

16. Yu M, Carter KT, Makar KW, Vickers K, Ulrich CM, Schoen RE, et al. Methylight droplet digital PCR for detection and absolute quantification of infrequently methylated alleles. Epigenetics. 2015;10:803–9.

17. Jensen SØ, Øgaard N, Ørntoft MBW, Rasmussen MH, Bramsen JB, Kristensen H, et al. Novel DNA methylation biomarkers show high sensitivity and specificity for blood-based detection of colorectal cancer- A clinical biomarker discovery and validation study. Clin Epigenetics. 2019;11.
18. Faaborg L, Jakobsen A, Waldstrøm M, Petersen CB, Andersen RF, Steffensen KD. HOXA9-methylated DNA as a diagnostic biomarker of ovarian malignancy. Biomark Med. 2021;15:1309–17.
19. Fung SYH, Chan KCA, Wong EWY, Ng CWK, Cho R, Yeung ZWC, et al. Droplet digital PCR of tumor suppressor gene methylation in serial oral rinses of patients with head and neck squamous cell carcinoma. Head Neck. 2021;43:1812–22.
20. Boeckx N, op de Beeck K, Beyens M, Deschoolmeester V, Hermans C, de Clercq P, et al. Mutation and Methylation Analysis of Circulating Tumor DNA Can Be Used for Follow-up of Metastatic Colorectal Cancer Patients. Clin Colorectal Cancer. 2018;17:e369–79.
21. Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Annals of Oncology. 2021;32:1167–77.
22. Chen X, Dong Z, Hubbell E, Kurtzman KN, Oxnard GR, Venn O, et al. Prognostic significance of blood-based multi-cancer detection in plasma cell-free DNA. Clinical Cancer Research. 2021;27:4221–9.
23. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden M V., Liu MC, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Annals of Oncology. 2020;31:745–59.
24. Fundamentals of Biostatistics. 2010.
25. Droplet Digital ™ PCR Droplet Digital ™ PCR Applications Guide.
26. Armbruster DA, Pry T. Limit of Blank, Limit of Detection and Limit of Quantitation. Clin Biochem Rev. 2008;29.
27. Digital D, Pcr ™. Rare Mutation Detection Best Practices Guidelines.
28. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016;54.
29. Sun XJ, Wang MC, Zhang FH, Kong X. An integrated analysis of genome-wide DNA methylation and gene expression data in hepatocellular carcinoma. FEBS Open Bio. 2018;8:1093–103.
30. Kim HJ, Kim N, Kim HW, Park JH, Shin CM, Lee DH. Promising aberrant DNA methylation marker to predict gastric cancer development in individuals with family history and long-term effects of H. pylori eradication on DNA methylation. Gastric Cancer. 2021;24:302–13.
31. Asada K, Nakajima T, Shimazu T, Yamamichi N, Maekita T, Yokoi C, et al. Demonstration of the usefulness of epigenetic cancer risk prediction by a multicentre prospective cohort study. Gut. 2015;64:388–96.

32. Chalasani NP, Ramasubramanian TS, Bhattacharya A, Olson MC, Edwards V DK, Roberts LR, et al. A Novel Blood-Based Panel of Methylated DNA and Protein Markers for Detection of Early-Stage Hepatocellular Carcinoma. Clinical Gastroenterology and Hepatology. 2021;19:2597-2605.e4.

33. Voutsadakis IA. Loss of genes in chromosome arms 5q and 16q in breast cancer. Breast Dis. 2022;41:331–41.

34. Acha P, Mallo M, Solé F. Myelodysplastic Syndromes with Isolated del(5q): Value of Molecular Alterations for Diagnostic and Prognostic Assessment. Cancers (Basel). 2022;14:5531.

35. Möller K, Kluth M, Ahmed M, Burkhardt L, Möller-Koop C, Büscheck F, et al. Chromosome 5 harbors two independent deletion hotspots at 5q13 and 5q21 that characterize biologically different subsets of aggressive prostate cancer. Int J Cancer. 2021;148:748–58.

36. Tanvir I, Hassan A, Albeladi F. DNA Methylation and Epigenetic Events Underlying Renal Cell Carcinomas. Cureus. 2022;

37. Faryna M, Konermann C, Aulmann S, Bermejo JL, Brugger M, Diederichs S, et al. Genome-wide methylation screen in low-grade breast cancer identifies novel epigenetically altered genes as potential biomarkers for tumor diagnosis. FASEB Journal. 2012;26:4937–50.

38. Hong SM, Omura N, Vincent A, Li A, Knight S, Yu J, et al. Genome-wide CpG island profiling of intraductal papillary mucinous neoplasms of the pancreas. Clinical Cancer Research. 2012;18:700–12.

39. Boeckx N, Op de Beeck K, Beyens M, Deschoolmeester V, Hermans C, De Clercq P, et al. Mutation and Methylation Analysis of Circulating Tumor DNA Can Be Used for Follow-up of Metastatic Colorectal Cancer Patients. Clin Colorectal Cancer. 2018;17:e369–79.

40. Jeanmougin M, Brodal HP, Dietrichson Pharo H, Vedeld HM, Lind GE. PoDCall: positive droplet calling and normalization of droplet digital PCR DNA methylation data. Bioinformatics. 2023;39.

41. Whale AS, Huggett JF, Tzonev S. Fundamentals of multiplexing with digital PCR. Biomol Detect Quantif. 2016;10:15–23.

42. Hawthorn L, Lan L, Mojica W. Evidence for field effect cancerization in colorectal cancer. Genomics. 2014;103:211–21.

43. Klouch KZ, Stern M-H, Trabelsi-Grati O, Kiavue N, Cabel L, Silveira AB, et al. Microsatellite instability detection in breast cancer using drop-off droplet digital PCR. Oncogene. 2022;41:5289–97.

44. Casas-Arozamena C, Moiola CP, Vilar A, Bouso M, Cueva J, Cabrera S, et al. Noninvasive detection of microsatellite instability in patients with endometrial cancer. Int J Cancer. 2023;

45. Favre L, Sako N, Tarfi S, Quang VT, Joy C, Dupuy A, et al. Evaluation of two new highly multiplexed PCR assays as an alternative to next-generation sequencing for IDH1/2 mutation detection. Mol Oncol. 2022;16:3916–26.

46. Visser E, de Kock R, Genet S, Borne B van den, Soud MY El, Belderbos H, et al. Up-front mutation detection in circulating tumor DNA by droplet digital PCR has added diagnostic value in lung cancer. Transl Oncol. 2023;27.

47. Oscorbin IP, Smertina MA, Pronyaeva KA, Voskoboev ME, Boyarskikh UA, Kechin AA, et al. Multiplex Droplet Digital PCR Assay for Detection of MET and HER2 Genes Amplification in Non-Small Cell Lung Cancer. Cancers (Basel). 2022;14.

48. Vannitamby A, Hendry S, Irving L, Steinfort D, Bozinovski S. Novel multiplex droplet digital PCR assay for scoring PD-L1 in non-small cell lung cancer biopsy specimens. Lung Cancer. 2019;134:233–7.

49. Ioannides M, Achilleos A, Kyriakou S, Kypri E, Loizides C, Tsangaras K, et al. Development of a new methylation-based fetal fraction estimation assay using multiplex ddPCR. Mol Genet Genomic Med. 2020;8.

50. Nyaruaba R, Li C, Mwaliko C, Mwau M, Odiwuor N, Muturi E, et al. Developing multiplex ddPCR assays for SARS-CoV-2 detection based on probe mix and amplitude based multiplexing. Expert Rev Mol Diagn. 2021;21:119–29.

51. Luo Y, Huang J, Tang Y, Luo X, Ge L, Sheng X, et al. Regional methylome profiling reveals dynamic epigenetic heterogeneity and convergent hypomethylation of stem cell quiescence-associated genes in breast cancer following neoadjuvant chemotherapy. Cell Biosci. 2019;9.

52. Zhao Y, O'Keefe CM, Hsieh K, Cope L, Joyce SC, Pisanic TR, et al. Multiplex Digital Methylation-Specific PCR for Noninvasive Screening of Lung Cancer. Advanced Science. 2023;10.

53. Manoochehri M, Wu Y, Giese NA, Strobel O, Kutschmann S, Haller F, et al. SST gene hypermethylation acts as a pan-cancer marker for pancreatic ductal adenocarcinoma and multiple other tumors: toward its use for blood-based diagnosis. Mol Oncol. 2020;14:1252–67.

Supplementary information

Supplementary Tables

Supplemental Table 1 | Amplification protocol.

Step	Temperature	Time	
Enzyme activation	95°C	10 min	
Denaturation	94°C	15s	**45X**
Amplification	Optimal T	1 min	
Inactivation step	98°C	10 min	
Hold	12°C	At least 15 min	

Supplemental Table 2 | Calculations

Sample size calculation*: Effect Size	$d = \frac{(M2\text{-}M1)}{\sqrt{\frac{(SD1^2 + SD2^2)}{2}}}$
Normalization of the number of droplets	$\frac{\text{Number of positive or negative droplets}}{\text{Number of accepted droplets}} * 20\,000$
Calculation of the methylation percentage	$\frac{\text{Normalized number of droplets target sequence}}{\text{Normalized number of droplets reference sequence}} * 100$
Limit of blank (LOB)	$\text{Mean number positive droplets}_{\text{blank}} + 1.645\ (SD_{\text{blank}})$
Limit of Detection (LoD)	$LOB + 1.645\ (SD_{\text{low sample concentration}})$
Calculation of the normalized number of droplets corrected with the LOB	$\frac{(\text{Number of positive or negative droplets} - LOB)}{\text{Number of accepted droplets}} * 20\,000$
Calculation of the methylation percentage corrected with the LOB	$\frac{\text{LOB corrected number of normalized droplets target sequence}}{\text{Normalized number of droplets reference sequence}} * 100$
Inter-assay variability coefficient (%CV)	$\frac{\text{Average}\ \frac{ng}{\mu L}\ \text{of triplicates}}{\text{Standard deviation}\ \frac{\text{conc}}{\mu L}\ \text{of triplicates}} * 100$
Detection sensitivity (%)	$\frac{Haploid\ genome\ equivalents\ (3)}{Number\ of\ copies\ per\ well} * 100$

* The sample size calculation is performed as described in Rosner B. Fundamentals of Biostatistics. 7th ed. Boston, MA: Brooks/Cole; 2011. For this, we need power, effect size and significance level. The power was set to 80%. The formula for the effect size is given in this table (M= mean, SD= standard deviation, 1= tumor group, 2= normal adjacent group). We chose the target with the smallest effect size (target 3, with an effect size of 2.21 in comparison to target 1 (2.31) and target 2 (2.8)). The significance level was 0.05. As there are three targets, we corrected for multiple testing using a one-sided test, since the targets are chosen to be hypermethylated in tumor, so we already know one group will have higher methylation levels.

Supplemental Table 3 | Overview of qPCR results

Target	**Fluorophore**	**Sample**	**AvCq**	**ΔCq**
1	FAM	Cell line	23,21	11,31
1	FAM	Blood	34,52	
2	FAM	Cell line	16,38	4,6
2	FAM	Blood	20,98	
2	SUN	Cell line	21,56	6,4
2	SUN	Blood	27,97	
3	FAM	Cell line	20,47	9,31
3	FAM	Blood	29,77	

Supplemental Table 4 | Comparison of the targets to *in silico* analyses of Ibrahim *et al.*

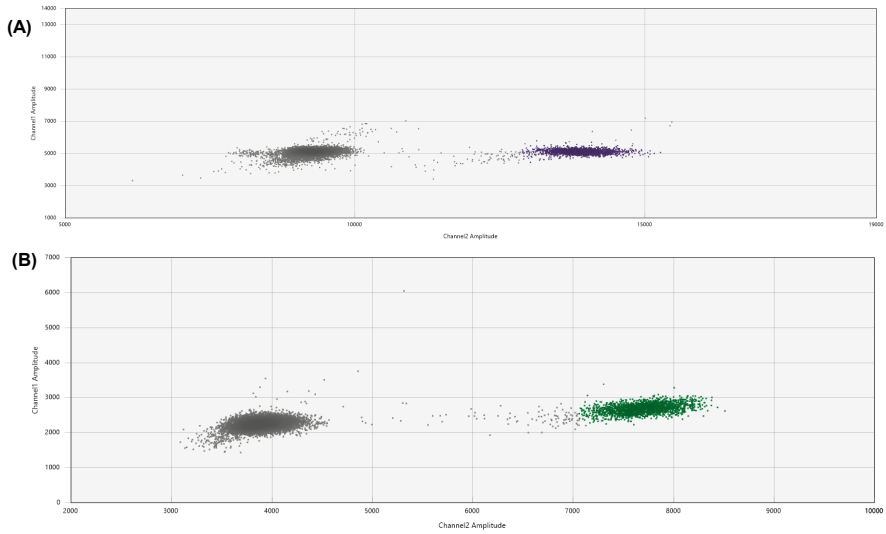
Target	*In silico* (cvAUC)	Wetlab (cvAUC)
1	0.855	0.909
2	0.894	0.876
1+2	0.933	0.949
3	0.870	0.847
1+2+3	0.947	0.948

cvAUC= cross-validated Area Under the Curve

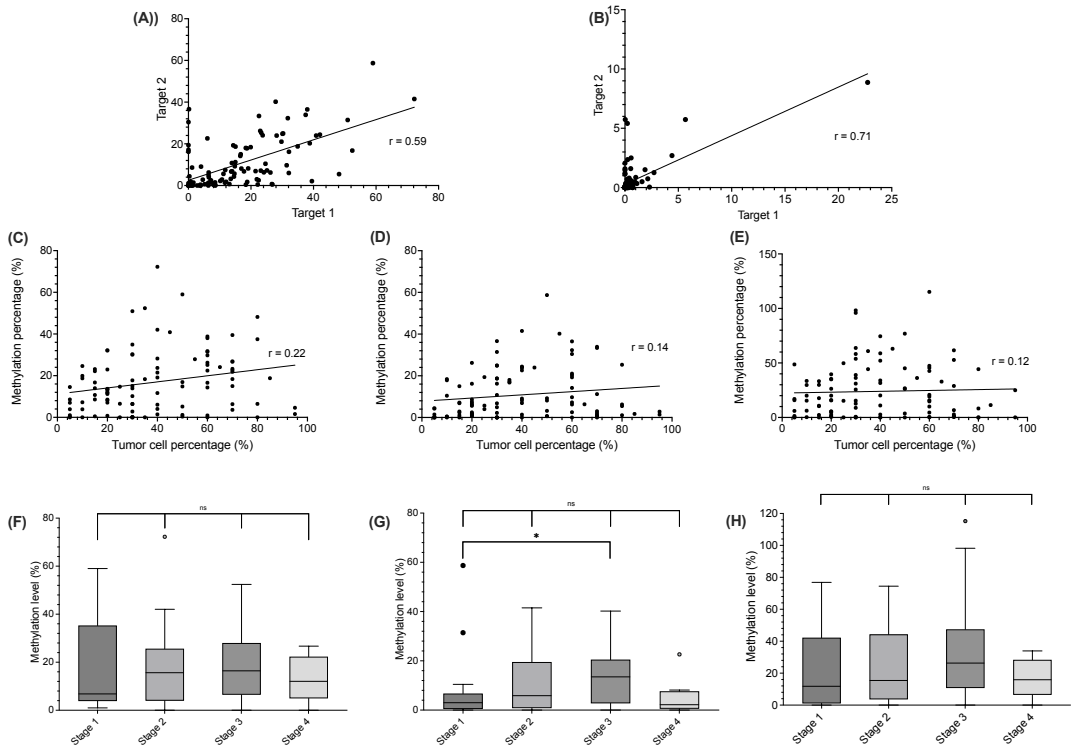
Supplementary Table 5| Sensitivity of targets per cancer stage (ROC analysis)

	Cancer Stage			
Target	**1** (n=16)	**2** (n= 28)	**3** (n= 27)	**4** (n= 9)
1	87,5%	89,9%	77,1%	88,9%
2	68,8%	77,8%	82,9%	66,7%
3	68,8%	77,8%	88,6%	88,9%
Total, 1/3	93,8%	100%	94,3%	88,9%
Total, 2/3	75,0%	74,1%	91,4%	88,9%

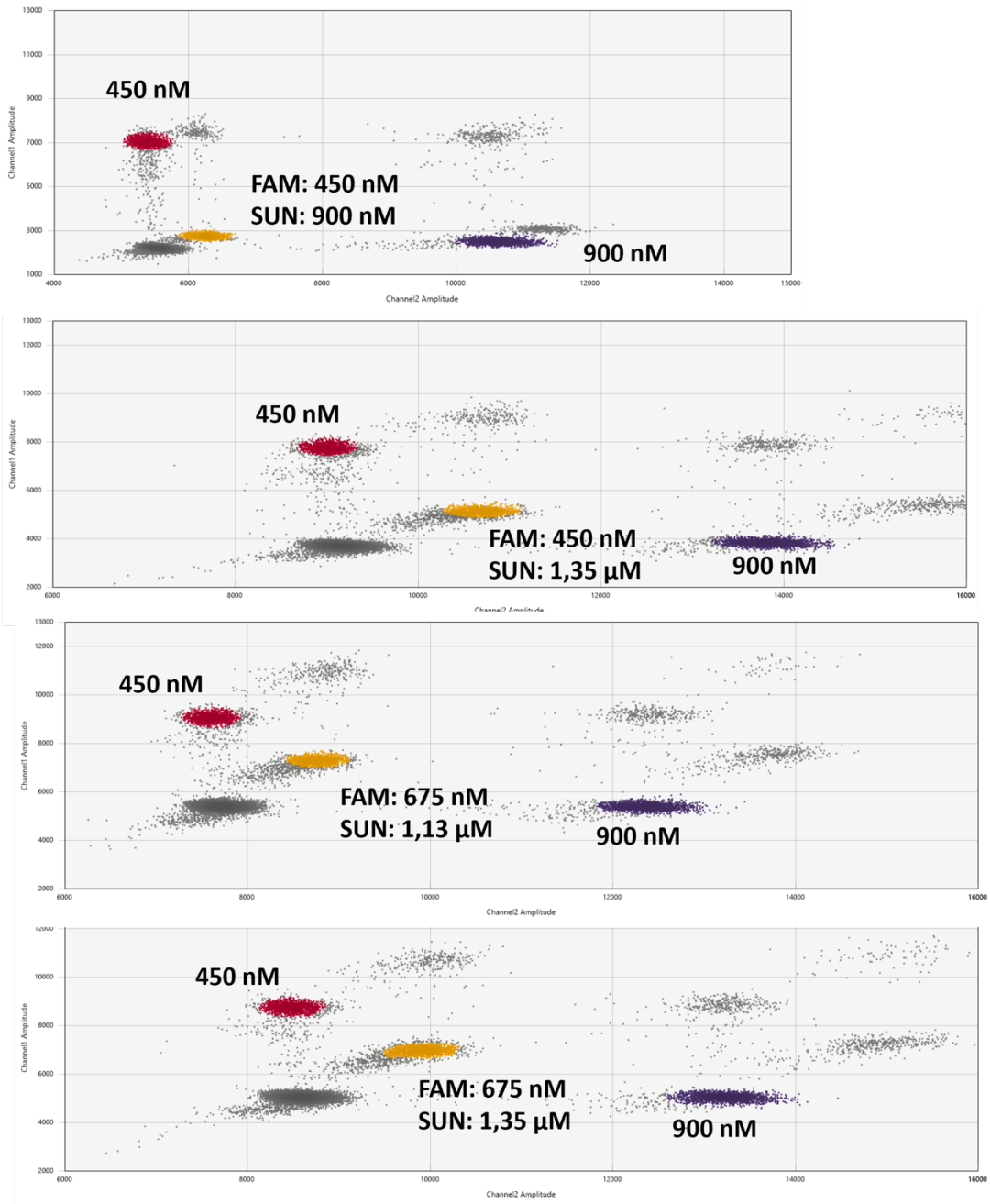
Supplementary Figures



Supplemental figure 1 | Dispersion graphs of negative samples in the triplex and duplex assay. The graphs show the dispersion of droplets for a negative control (gDNA from a whole blood sample) A) Triplex assay consisting of target 1, 2 and reference Albumin. B) Duplex assay consisting of target 3 and reference albumin.



Supplemental figure 2 | Correlations of methylation levels per target. A-B) Methylation levels in the triplex assay. C-E) Methylation levels of samples and tumor cell percentage (TcP). F-H) Methylation levels per cancer stage. Kruskal-Wallis tests and post-hoc Mann-Whitney U tests were used. Ns = not significant, *= p-value < 0.05. Analyses and plotting were performed using GraphPad Prism.



Supplemental figure 3 | Different probe concentrations for cluster separation. This figure illustrates the increased cluster separation at different probe concentrations for the triplex and the duplex assay. Probe concentrations are given in the figure.

CHAPTER 6

Adapted from

IMPRESS: Improved Methylation Profiling using Restriction Enzymes and smMIP Sequencing, Combined with a New Biomarker Panel, Creating a Multi-Cancer Detection Assay

(Manuscript accepted at the *British Journal of Cancer*)

Janah Vandenhoeck*, Isabelle Neefs[1,2*], Thomas Vanpoucke[1,2*], Joe Ibrahim[1,2], Arvid Suls[1,2], Dieter Peeters[3], Anne Schepers[1], Alexander Hoischen[4,5], Erik Fransen[1], Marc Peeters[2], Guy Van Camp[1,2*] and Ken Op de Beeck[1,2*]

**Janah Vandenhoeck, Isabelle Neefs and Thomas Vanpoucke have contributed equally to this work (shared first authors), as well as Guy Van Camp and Ken Op de Beeck (shared last authors)*

1 Centre of Medical Genetics, University of Antwerp and Antwerp University Hospital, Prins Boudewijnlaan 43, 2650 Edegem, Belgium

2 Centre for Oncological Research Antwerp (CORE), University of Antwerp and Antwerp University Hospital, Universiteitsplein 1, 2610 Wilrijk, Belgium

3 Department of Pathology, Antwerp University Hospital, Drie Eikenstraat 655, 2650 Edegem, Belgium

4 Department of Human Genetics and Radboud Institute for Molecular Life Sciences, Radboud University Medical Centre, Geert Grooteplein Zuid 10, 6525 GA Nijmegen, The Netherlands

4 Department of Internal Medicine and Radboud Institute for Infectious Diseases (RCI), Radboud University Medical Centre, Geert Grooteplein Zuid 10, 6525 GA Nijmegen, The Netherlands

ABSTRACT

Background: Despite the worldwide progress in cancer diagnostics, more sensitive diagnostic biomarkers are needed. The methylome has been extensively investigated in the last decades, but a low-cost, bisulfite-free detection method for multiplex analysis is still lacking.

Methods: We developed a methylation detection technique called IMPRESS, which combines methylation-sensitive restriction enzymes and single-molecule Molecular Inversion Probes. We used this technique for the development of a multi-cancer detection assay for eight of the most lethal cancer types worldwide. We selected 1,791 CpG sites that can distinguish tumor from normal tissue based on DNA methylation. These sites were analyzed with IMPRESS in 35 blood, 111 tumor and 114 normal samples. Finally, a classifier model was built.

Results: We present the successful development of IMPRESS. The final classifier model discriminating tumor from normal samples was built with 358 CpG target sites and reached a sensitivity of 0.95 and a specificity of 0.91. Moreover, we provide data that highlights IMPRESS's potential for liquid biopsies.

Conclusions: We successfully created an innovative DNA methylation detection technique. By combining this method with a new multi-cancer biomarker panel, we developed a sensitive and specific multi-cancer assay, with potential use in liquid biopsies.

INTRODUCTION

Cancer remains one of the most lethal diseases worldwide. Breast, lung, colorectal, and prostate cancer are amongst the most common cancers, with each over 1.4 million cases per year [1]. Diagnosis typically occurs in an advanced disease stage due to the lack of clear symptoms and the absence of effective screening programs for most cancer types [2]. This is reflected in the percentages of late-stage diagnoses, for example, 68% and 59% in lung and colorectal cancer respectively [3]. Clearly, there is an unfulfilled need for effective diagnostic biomarkers.

An interesting biomarker candidate for cancer detection is DNA methylation. Genome wide epigenetic reprogramming of tumors occurs early in carcinogenesis. Methylation patterns of many tumor types are widely dysregulated compared to those of healthy cells, but the tumor type specific patterns are very distinctive [4–6]. Many studies have investigated the potential of the methylome in recent years but, to date, there are only a few successful methylation biomarkers for cancer in a clinical setting. Our research group has already shown the promise of methylation as a diagnostic biomarker [6]. We demonstrated the capability to discriminate 14 different cancer types from normal tissue and from each other using methylation biomarkers *in silico,* with high sensitivity and specificity.

It is important to detect biomarkers in a sensitive and specific manner. Currently, bisulfite sequencing is still considered the gold standard for DNA methylation analysis. However, bisulfite is a harsh chemical that degrades DNA, limiting the sensitivity of downstream applications [7]. Alternative bisulfite-free techniques, such as affinity-based methods (e.g. MeDIP-seq, Methyl Cap), are shown to have a lower accuracy compared to bisulfite sequencing [8,9]. Other bisulfite-free methods include restriction enzymes, which can either be methylation-dependent, digesting only methylated CpGs, (e.g. MspJI) or methylation-sensitive restriction enzymes, digesting only unmethylated CpGs (e.g. HpaII) [10].

Recently, two enzymatic technologies for DNA methylation detection were launched, called TET-assisted pyridine borane sequencing (TAPS) and enzymatic methyl sequencing (EM-Seq) [11,12]. In TAPS, ten-eleven translocation (TET) oxidation is combined with pyridine borane reduction [12]. EM-seq consists of two conversion steps as well, using TET2 and APOBEC3A [11,13] The enzymatic treatment in these techniques is a first step towards eliminating the need for bisulfite [11,13,14]. However, genome-wide methylation detection techniques come at a high cost which hampers their implementation in routine diagnostics [15]. Clearly, there is an urgent need for a low-cost, bisulfite-free detection method for the simultaneous analysis of multiple methylation regions that allows accurate prediction of disease.

In this paper, we describe a novel high-multiplex methylation detection technique called IMPRESS (Improved Methylation Profiling using Restriction Enzymes and smMIP sequencing). Methylation-sensitive restriction

enzymes (MSREs) have already been used for a very long time for the analysis of methylation in specific regions of the genome [9]. Single-molecule Molecular Inversion Probes (smMIPs) are very efficient for capturing and enriching carefully chosen informative regions of the genome. They are extremely suitable for multiplex analysis of thousands of genomic regions [16]. smMIPs have been used for mutation detection and CNV analyses in different research fields [17,18] but have never been described for DNA methylation detection. We used this technique for the development of a diagnostic biomarker assay discriminating tumor samples from normal samples. For the selection of the biomarker targets, methylation data of tissue samples from eight of the most lethal cancer types worldwide and normal adjacent tissue were used. Furthermore, methylation patterns of normal blood samples were included for the target selection, ensuring the development of a biomarker panel that is also suitable for use in plasma derived liquid biopsies in the future. In this study, the proof-of-principle of the biomarker assay combining the new IMPRESS technique and the biomarker panel, is described.

MATERIALS AND METHODS

Development of new DNA methylation detection technique

The IMPRESS technique is a novel combination of MSRE digestion and smMIP sequencing. An overview of the technique is given in Fig. 1 and the molecular details are shown in Suppl. Fig. 1. Details of the protocol are described in the Suppl. Methods.

IMPRESS protocol

In brief, the first step was a combined digestion of 50ng DNA with four MSREs. The MSREs cleave unmethylated DNA at their recognition sites, while methylated CpG sites block the restriction enzymes, which results in unaffected CpG regions. The methylated CpG regions were captured by the smMIPs through hybridization of the smMIP binding arms. Elongation and ligation of the smMIP created a circular DNA fragment. All remaining linear fragments were degraded by an exonuclease reaction. Thereafter, the circular fragments were amplified by PCR. Finally, all samples were pooled, purified, and sequenced by Next Generation Sequencing. In each sample, lambda phage DNA was spiked in as an internal digestion control.

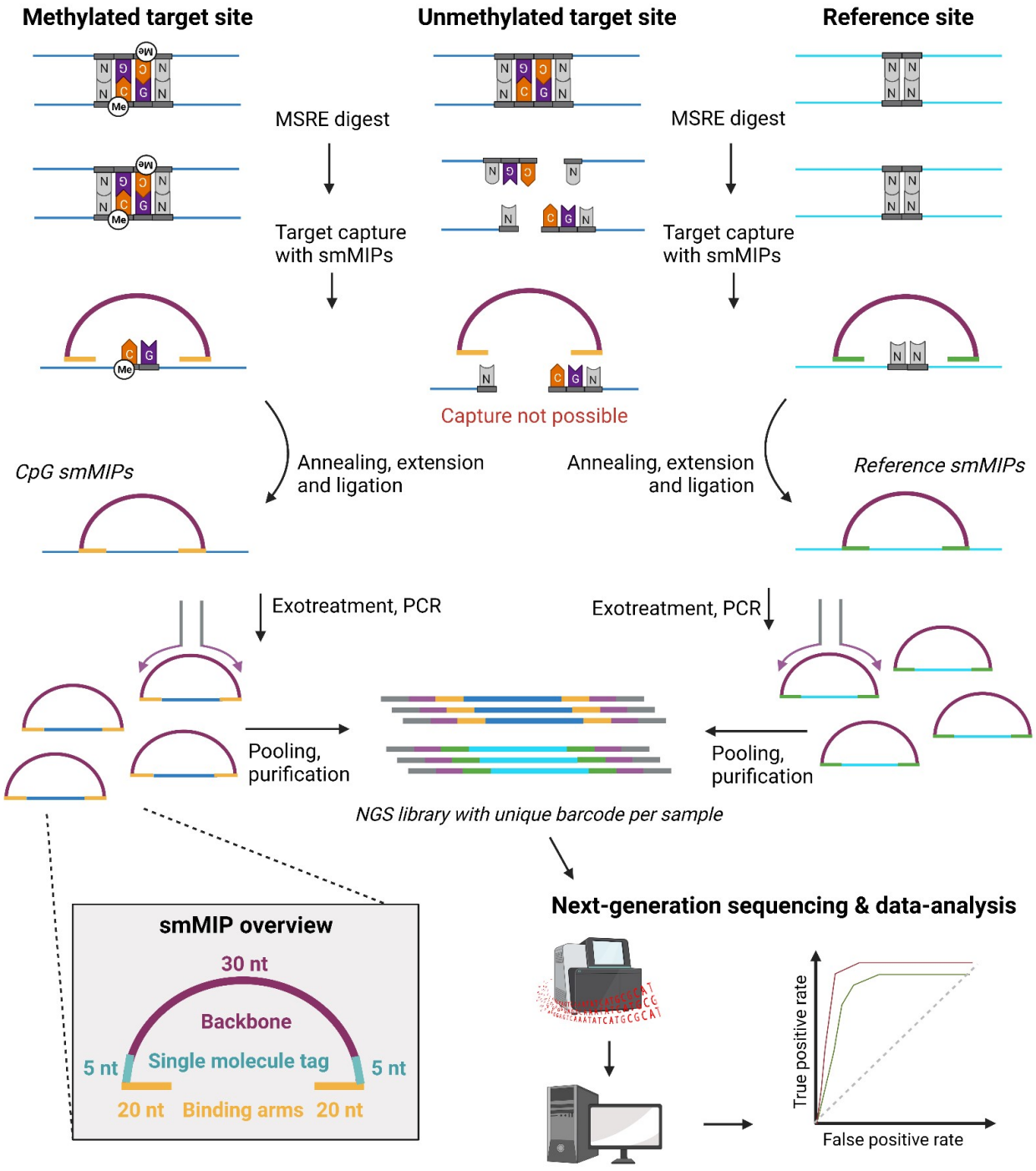


Figure 1 | Overview of the IMPRESS technology. During combined digestion with four MSREs, unmethylated recognition sites are digested, whereas methylated CpG sites, blocking the MSREs, remain intact. Subsequently, methylated CpG sites are captured by hybridization of the smMIP binding arms. Both the selected CpG sites and reference sites are targeted by specifically designed smMIPs, referred to as CpG smMIPs and reference smMIPs, respectively. The insert gap is extended and ligated to form a circular fragment. All linear fragments are degraded by an exonuclease treatment and the circular fragments are amplified by PCR with universal primers binding the common smMIP backbone. Eventually, PCR products are pooled, purified and sequenced simultaneously. Created with BioRender.com

Data analysis pipeline

For the analysis of the NGS output, a bioinformatic pipeline was built using *Snakemake* [19]. Configuration was done in a *json* file, where all pipeline parts can be configured separately. Both MiSeq and NextSeq output can be handled, and as the pipeline has a modular structure, parts can be added or removed very easily. Computational parallelization was achieved per sample for both MiSeq and NextSeq data.

First, the reads were split per sample by using the two sample barcode reads (Suppl. Fig. 1), allowing for one base error per barcode. Reads were then quality trimmed and subsequently mapped to either the human (hg19) or lambda phage genome using the *bwamem* algorithm with default parameters [20]. Next, *Picard MarkDuplicates* was used for duplicate removal, based on the single molecule tags of ten nucleotides in total [21]. Then, reads were filtered out based on mapping quality and mapping flags, keeping only properly paired reads with quality above 15. Finally, reads were counted per smMIP location, only retaining a pair when both reads match the target location within an error margin of 5 base pairs. This way, a dataset with counts for all smMIP.

Code availability

Code, including all version information, is available upon request as a code ocean capsule, where the code can be run in an online environment or downloaded (https://doi.org/10.24433/CO.5112387.v1).s for each sample was obtained.

Internal control for digestion

To check whether each sample was sufficiently digested, we used the read counts of the smMIPs targeting spiked-in lambda phage DNA. Two types of lambda phage DNA smMIPs were included: (a) smMIPs targeting a CpG with one of the MSRE recognition sites, and (b) smMIPs targeting a reference site, without recognition site or without CpG site. The percentage of non-digested DNA of each sample was calculated as described below. A threshold of 5% non-digested fragments was set.

$$Percentage\ of\ non-digestion_{digested\ sample} =$$

$$\frac{\left(\frac{\sum CpG\ \ smMIP\ count_{digested\ sample}}{\sum Reference\ \ smMIP\ count_{digested\ sample}}\right)}{Average_{all\ undigested\ samples}\left(\frac{\sum CpG\ \ smMIP\ count_{undigested\ sample}}{\sum Reference\ \ smMIP\ count_{undigested\ sample}}\right)}$$

Development of multi-cancer biomarker panel

Target sites selection

For the development of the multi-cancer detection assay, a panel with candidate methylation biomarkers was built using online available 450K methylation array data (Table 1). Methylation data processing and analysis were performed based on the methods previously described by Ibrahim *et al.*[6].

smMIP design

Using the *MIPGEN* software [22], smMIPs were designed for both DNA strands (i.e. double-tiled) for each selected target site. smMIPs in our design contain (a) a common smMIP backbone of 30nt, (b) single molecule tags of 5nt on each side, and (c) two binding arms of circa 20nt that were specifically designed for each target to have an insert size of 50nt (Fig. 1). The single molecule tags differ per smMIP copy and allow filtering for PCR duplicates. Next, smMIPs covering SNPs and/or repeats were removed, and a final selection was made.

Sample collection and processing

Control samples

Lambda phage DNA was purchased from Thermo Fisher Scientific (USA). Human methylated and non-methylated (WGA) DNA was purchased from Zymo Research (USA). Four cancer cell lines were provided by the Centre for Oncological Research (CORE, Antwerp), and one line was purchased from the German Collection of Microorganism and Cell cultures (DSMZ, Germany) (Suppl. Table 1). All cell lines were cultured according to standard protocols from the American Type Culture Collection (ATCC). The cell lines were authenticated at the start of the study and routinely tested for mycoplasma contamination, which was negative. Genomic DNA was extracted using the Blood & Cell culture DNA Maxi kit (Qiagen, Germany). DNA was stored at -20°C until further use. For the liquid biopsy experiments, cfDNA material was provided by the diagnostics department of the Center of Medical Genetics. This cfDNA was anonymized residual material obtained from NIPT plasma samples. cfDNA was stored at -20°C until further use. For limited research use of this type of residual material, no additional ethical approval is required.

Blood samples

A total of 35 whole blood samples were collected from healthy volunteers. Genomic DNA (gDNA) was extracted using a standard salting-out process. The DNA was stored at 4°C until further use.

Fresh frozen tissue

Tumor tissue and normal adjacent tissue samples were routinely collected by the biobank of the Antwerp University Hospital (UZA, Belgium). All normal tissue cited throughout this manuscript is normal tissue adjacent to tumor tissue, except for the normal breast samples, which originate from breast reductions of healthy women. A total of 225 fresh frozen tissues stored at -80°C were used (Table 1). Tissue type, presence of invasive tumor, and overall tumor cell percentage (TcP) were verified by a pathologist (D.P.) through microscopic examination of hematoxylin-eosin-stained sections. Samples with a minimum of 5% TcP were retained for analysis. DNA was extracted using the QIAamp DNA Micro Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The DNA was stored at -20°C until further use.

Table 1 | Amount of samples used for target selection and sequencing experiments.

		Repeatability		**Selection of CpG sites***		**Multi-cancer assay°**	
(Tumor) tissue type	**Mortality rank**	**# Tumor**	**# Normal**	**# Tumor**	**# Normal**	**# Tumor**	**# Normal**
Lung cancer (LUAD + LUSC)	1	3	2	370 + 473	42 + 32	9 + 13	11 + 11
Colorectal adenocarcinoma (CRC)	2	3	2	411	45	7	10
Liver hepatocellular carcinoma (LIHC)	4	2	2	377	50	13	11
Breast invasive carcinoma (BRCA)	5	2	0	791	96	10	9
Pancreatic adenocarcinoma (PAAD)	6	2	2	184	10	21	24
Head & neck squamous cell carcinoma (HNSC)	7	2	1	528	50	13	8
Esophageal carcinoma (ESCA)	8	2	1	185	16	10	5
Prostate adenocarcinoma (PRAD)	9	2	1	502	50	15	25
Blood	-	-	-	-	140	-	35
Total		**18**	**11**	**3,821**	**531**	**111**	**149**

Stomach cancer ranks 3rd in mortality but was left out of the analysis due to a lack of normal samples in the public dataset at the time of analysis. LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma.

*CpG sites were selected from online available 450K methylation array datasets. Tumor and normal tissue sample data were retrieved from The Cancer Genome Atlas (TCGA). Blood datasets, retrieved from the Gene Expression Omnibus (GEO), originate from 18 peripheral blood mononuclear cells (PBMC) (GSE111942), 4 left atrium (GSE62727), 30 erythrocyte progenitors from bone marrow (GSE63409), 6 whole blood, 6 PBMCs, 6 granulocytes, 6 CD4+ T cells, 6 CD8+ T cells, 6 CD56+ NK cells, 6 CD19+ B cells, 6 CD14+ monocytes, 6 neutrophils and 6 eosinophils (GSE35069) and 28 peripheral blood (GSE113012).

°For the multi-cancer assay, we used both matched and unmatched tissue samples. Matched samples are tumor and adjacant normal tissue samples from the same patient. We used 8 matched samples for LUAD, 10 for LUSC, 6 for CRC, 7 for LIHC, 4 for BRCA, 7 for PAAD, 5 for HNSC, 2 for ESCA, and 10 for PRAD. Fot the ddPCR assay, we used most of these samples again.

Classifier model construction based on NGS data

For each smMIP, linear discriminant analysis was carried out using the *lda* function from the *MASS* package in the software package R (version 4.0.2) [23]. A model was first constructed and then validated using the *ROCR* package [24]. Five-fold cross validation was carried out with a randomization restriction to proportionally represent the tumor types across the five folds. Predictive accuracy of the LDA models was expressed using the Area under the ROC curve (AUC).

Finally, the least efficient smMIPs were removed with a cutoff of 1,000 cumulative counts in all undigested samples, since a minimal number of counts is needed to make a robust classifier. All smMIP models with a cross-validated AUC below 0.8 were removed for the final model. In case of double-tiled smMIPs, the best performing smMIP was selected. All remaining single smMIP models were combined into the final model. The prediction cutoff for each single smMIP model was determined by the lowest sum of false positives and false negatives. The combination of all single predictions was then assessed by a ROC curve and the prediction cutoff was determined based on the highest overall accuracy to produce the final classifier model.

Statistics and calculations

For the power calculations, we used online datasets to obtain the mean and the standard deviation of the methylation level of the selected targets for the tumor and the normal group. We used the target with the smallest Cohen's D effect size. A sample size with 67 cases and an equal number of controls holds 80% power to detect any difference between the tumor group (methylation level = 0.50 ± 0.24) and the normal group (methylation level = 0.30 ± 0.20), corrected for multiple testing (1791 CpG sites) with a two-sided test. We were able to collect 111 tumor samples and 149 normal samples. This holds a power of 99%.

For the statistical analyses, differences in average methylation levels between tumor and normal adjacent samples within one tissue type were tested using the Mann-Whitney U test (two-sided). The performance of the IMPRESS was expressed in terms of specificity and sensitivity. To measure the repeatability of our technique, the Pearson correlation between normalized counts from two separate sequencing runs calculated. In addition, a Bland-Altman analysis was performed using the normalized counts of two independent runs.

For all analyses, p-values lower than 0.05 were considered significant. All statistical tests were performed in R (version 4.0.2) or GraphPad Prism (version 10.0.0) for macOS, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com.

Normalized counts for each smMIP were calculated as follows:

$$Normalized\ count\ smMIP\ i\ in\ sample\ A = \frac{Absolute\ count\ i\ in\ A}{\sum reference\ smMIP\ counts\ \ in\ A}$$

Sensitivity, specificity, accuracy and balanced accuracy are calculated as follows with the following abbreviations: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Balanced\ accuracy = \frac{Sensitivity + Specificity}{2}$$

Ethical approval

This study was conducted in accordance with Good Clinical Practice guidelines and the Declaration of Helsinki. Fresh frozen tissue samples used in this study were previously collected in the Biobank of the Antwerp University Hospital and retrospectively used in this study. According to Article 20 of the Belgian Law on the procurement and use of human corporal material intended for human application or scientific research of 19 December 2008, patients provide consent for the use of their bodily material in research when consenting to an invasive procedure. As such, no additional consent was needed for the use of these retrospective samples. For prospectively collected blood samples, informed consent was given by each subject. The study protocol and any modifications thereof were approved by the UZA ethical committee (Ref. N°20/02/056 and Ref. N°41/14/426) before experimental analyses were performed.

RESULTS

Development of novel DNA methylation detection technique

IMPRESS technique

The first part of this study was the development and optimization of the IMPRESS technique, which combines MSREs and smMIPs [16,25,26]. This unique combination allows the use of smMIPs for DNA methylation detection, which has never been described before. An overview of the technique is given in Fig. 1.

The first step was a combined digestion of the DNA with four MSREs (HpaII, HpyCH4IV, AciI and HinP1I). Each MSRE recognition site has a CpG site in the middle (C^CGG, A^CGT, C^CGC and G^CGC). The efficiency of the four MSREs is optimal in the same buffer and at the same temperature, and together they cover 39% of all CpG sites in the human genome [9]. During digestion, unmethylated recognition sites were cleaved. Methylated CpG sites blocked the restriction enzymes which resulted in uncut CpG sites. As a control, undigested samples were also included and treated similarly, except for the omission of MSREs. As a control for digestion, lambda phage DNA was spiked into each sample. After the combined digest, all recognition sites in the lambda DNA were expected to be cleaved, as lambda DNA is not methylated.

In the next step, CpG sites of interest were targeted by a pool of phosphorylated smMIPs. A smMIP is a DNA fragment containing a common backbone of 30nt, single molecule tags of 5nt on each side, and two binding arms of circa 20nt (Fig. 1). Unmethylated CpG regions were cleaved by the MSREs and therefore smMIP capturing was not possible in these regions. Methylated CpG regions remained intact and were captured by the hybridization of the binding arms of the smMIPs. In addition, regions without enzyme recognition sites were targeted as a reference. Elongation and ligation of the smMIP created a circular DNA fragment. In all capture reactions, some smMIPs were ligated without a 50nt insert. These so-called empty smMIPs were eventually removed through purification and through filtering during data analysis.

After the capture reaction, an exonuclease treatment was performed, in which all linear fragments such as unbound smMIPs or original DNA strands were degraded, and only circular fragments remained intact. These fragments were amplified by PCR. Finally, all fragments were purified and sequenced by NGS.

In addition to our wet lab protocol, an accompanying bioinformatics analysis pipeline was developed to process the NGS data. Using a *Snakemake* workflow, all sequencing reads were deconvoluted per sample and mapped to the genome. Next, all duplicates were removed. After quality filtering, reads per smMIP location were counted for each sample and a counts table was obtained.

Efficiency of the methylation-sensitive restriction enzymes

To test the cutting efficiency of the combination of the four selected MSREs, both a methylated and an unmethylated lambda DNA sample were digested by the MSREs. Subsequently, the digested samples were amplified in triplicate with primer pairs hybridizing around one of the MSRE recognition sites in a qPCR experiment. Undigested, (un)methylated lambda DNA samples were also included as positive controls.

According to the Lightcycler software (Roche), the undigested samples had an average Ct value of 5.0 and the unmethylated digested sample had an average Ct value of 22.0 (ΔCt=17) (Fig. 2). This means that merely 1 in 2^{17}=~131,000 DNA molecules were not digested. Thus, the remaining fraction of undigested DNA in unmethylated samples is negligible. The methylated digested sample had an average Ct value of 5.7, indicating that methylation effectively blocks digestion by the MSREs.

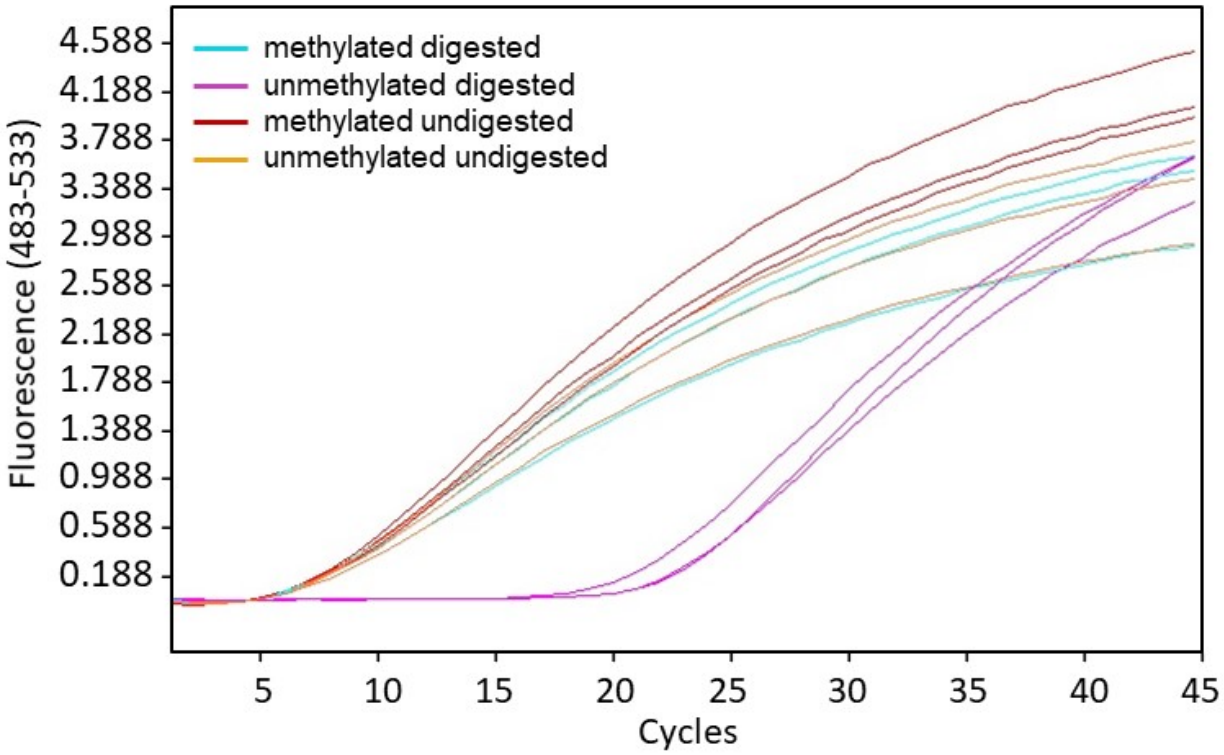


Figure 2 | Efficiency of the methylation-sensitive restriction enzymes. Amplification curve of qPCR of (un)methylated and (un)digested lambda DNA. This is an example of one primer pair hybridizing around a MSRE recognition site. Two types of control samples were used, unmethylated lambda DNA and artificially methylated lambda DNA by CpG methyltransferase M.SssI (New England Biolabs). Each sample was amplified in triplicate. Methylated digested (blue), unmethylated digested (purple), methylated undigested (red) and unmethylated undigested (orange) samples have an average Ct value of 5.7, 22.0, 5.0 and 5.0, respectively.

Repeatability

To evaluate the repeatability of the IMPRESS technique, two independent experiments were performed on a set of 29 fresh frozen tissue samples (Table 1). We used the smMIPs designed for the multi-cancer biomarker

panel (see Development of a multi-cancer biomarker panel). The two libraries were sequenced together, and data processing was performed. The normalized counts (calculation in Statistics and calculations) of the same samples for both experiments were compared, and a Pearson correlation coefficient of r=0.99 was obtained (Fig. 3a). In the Bland-Altman analysis, the bias was 0.007502 ± 0.03066 with 95% limits of agreement of -0.05259 and 0.06759 (Fig. 3b). These results prove the high repeatability of the IMPRESS technique.

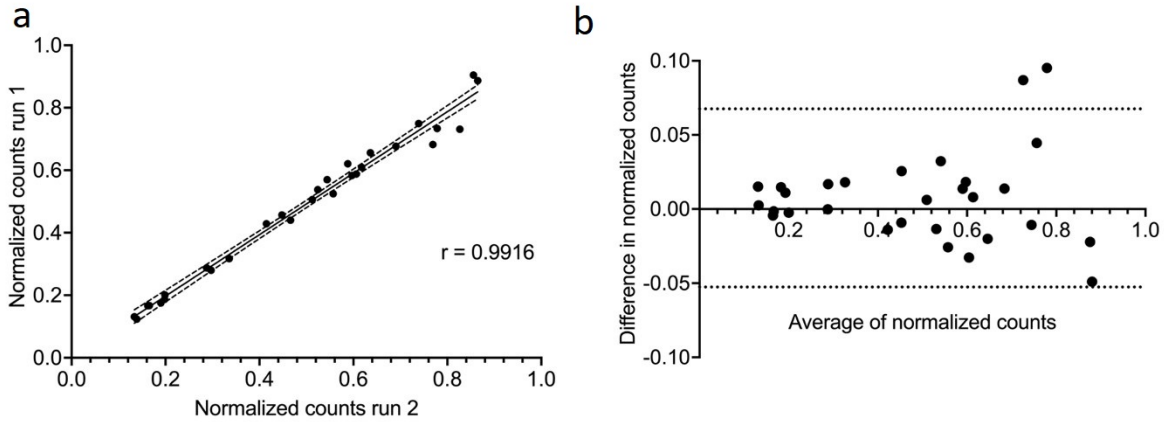


Figure 3 | Repeatability of novel IMPRESS technology. Repeatability of the technique by comparing two independent experiments with the same samples. (a) The normalized counts of the samples have a correlation coefficient of r=0.99. (b) Bland-Altman comparison for the repeatability. Bias is 0.007502 ± 0.03066 with 95% limits of agreement [-0.05259; 0.06759]. Plotting was performed using GraphPad Prism.

Development of a multi-cancer biomarker panel

The second part of this study was the development of a multi-cancer diagnostic biomarker panel, which could subsequently be validated by the IMPRESS technique to develop a multi-cancer detection assay. For the selection of potential biomarkers, online available methylation data of eight of the most lethal cancer types worldwide and healthy blood samples were used (Table 1). In total, 1,791 hypermethylated CpG sites (Fig. 4) were selected based on the following parameters: (a) a minimum average methylation level of 0.5 in the tumor tissue samples of eight cancer types, (b) a maximum average methylation level of 0.3 for the normal adjacent tissue and blood, and (c) the presence of at least one restriction site for one of the four used MSREs. On average, four sites were interrogated per region. Normal blood datasets were included for biomarker selection, resulting in a biomarker panel that is suitable for liquid biopsies as well. Secondly, a total of 2,739 reference sites were selected from the human genome. These reference sites were included to estimate the effective total amount of input DNA and allow normalization of the results per sample. Reference sites were chosen to (a) not contain a recognition site of the selected MSREs (1,000 sites), or (b) not contain a CpG site (1,739 sites). These regions are never cleaved by the enzymes and are therefore always captured by the smMIPs. Lastly, both CpG (10 per MSRE site) and reference sites (15 without recognition site, 15 without CpG site) were selected in lambda phage

DNA. Lambda phage DNA is never methylated and is used as an internal control for the enzymatic digestion reaction.

For these 1,791 hypermethylated CpG sites, 2,739 reference sites, and 70 lambda phage DNA sites, smMIPs were designed for both DNA strands (Fig. 4). After the removal of smMIPs covering SNPs and/or repeats, 2,331 CpG smMIPs and 600 reference smMIPs (300 without recognition site and 300 without CpG site) were selected for human targets. For lambda phage DNA, 12 CpG smMIPs (3 per restriction site) and 10 reference smMIPs (5 without recognition site and 5 without CpG site) were selected.

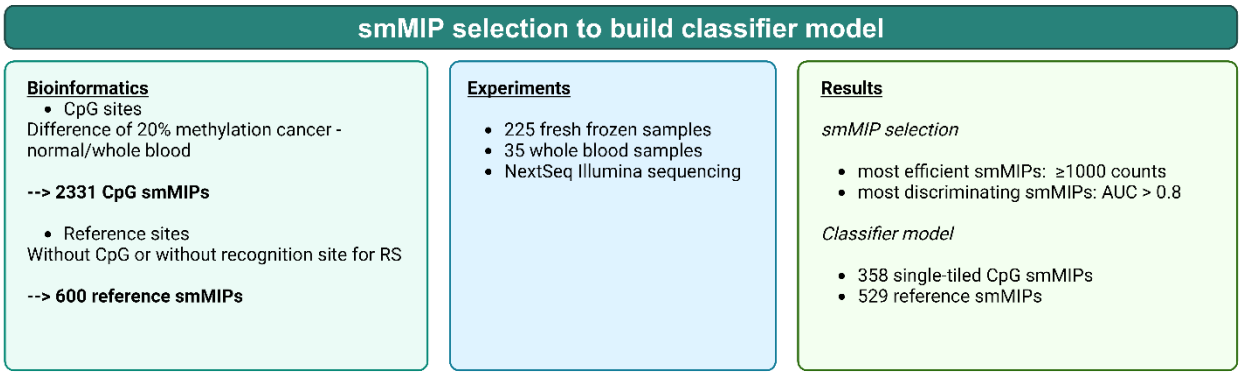


Figure 4 | Overview of the main analyses for the multi-cancer detection assay. The main experiment is shown: The experiment is used for the determination of the final classifier model. Created with BioRender.com.

Multi-cancer detection assay

Data exploration

To evaluate the biomarker panel and the IMPRESS technique, we performed an experiment on fresh frozen tissue and blood samples (Fig. 4). First, we prepared a sequencing library with 111 tumor samples, 114 normal adjacent tissue samples and 35 whole blood samples (Table 1) targeted by a total of 2,953 smMIPs (2,331 CpG smMIPs, 600 reference smMIPs and 22 lambda smMIPs). Capillary electrophoresis analysis of this library is shown in Suppl. Fig. 2. Sequencing was performed on the Illumina NextSeq system and NGS parameters are shown in Suppl. Table 2. After the first analysis of the raw data, read counts for all smMIPs for each sample were obtained. Based on the characterization experiments, a minimum read count threshold of 5000 counts per sample was determined, and all samples met this requirement.

The efficiency of the MSRE digest was verified in each sample by the spiked-in lambda phage DNA as an internal control (calculations in part 2.1.3). A threshold of 5% non-digested fragments was set. One out of 260 samples exceeded this threshold (9.7%) and was removed from further analyses. On average, only 1.3% of the DNA in each sample was not properly digested (Suppl. Fig. 3). In total, 19 underperforming CpG smMIPs with no counts in any sample were removed from the analysis. Finally, normalization was executed per sample to correct for

the amount of input DNA. Hereto, we divided the counts of every CpG smMIP by the sum of all reference smMIP counts, resulting in a final dataset with counts of 2,312 CpG targeting smMIPs for 259 samples. The normalized count is assumed to be higher in samples methylated for our targets (i.e. tumor samples) and lower in samples unmethylated for our targets (i.e. normal samples). An overview of the sample distribution of the sum of the normalized CpG smMIP counts is given in Fig. 5. Tumor samples showed higher and more spread normalized counts, while normal samples showed lower and more similar normalized counts. However, the normalized counts for normal colorectal tissue are higher than all other normal tissues and overlap with some of cancer types. The blood samples had the lowest normalized counts compared to all other tissue types. Within each tissue type, the average normalized counts of tumor and normal samples were significantly different.

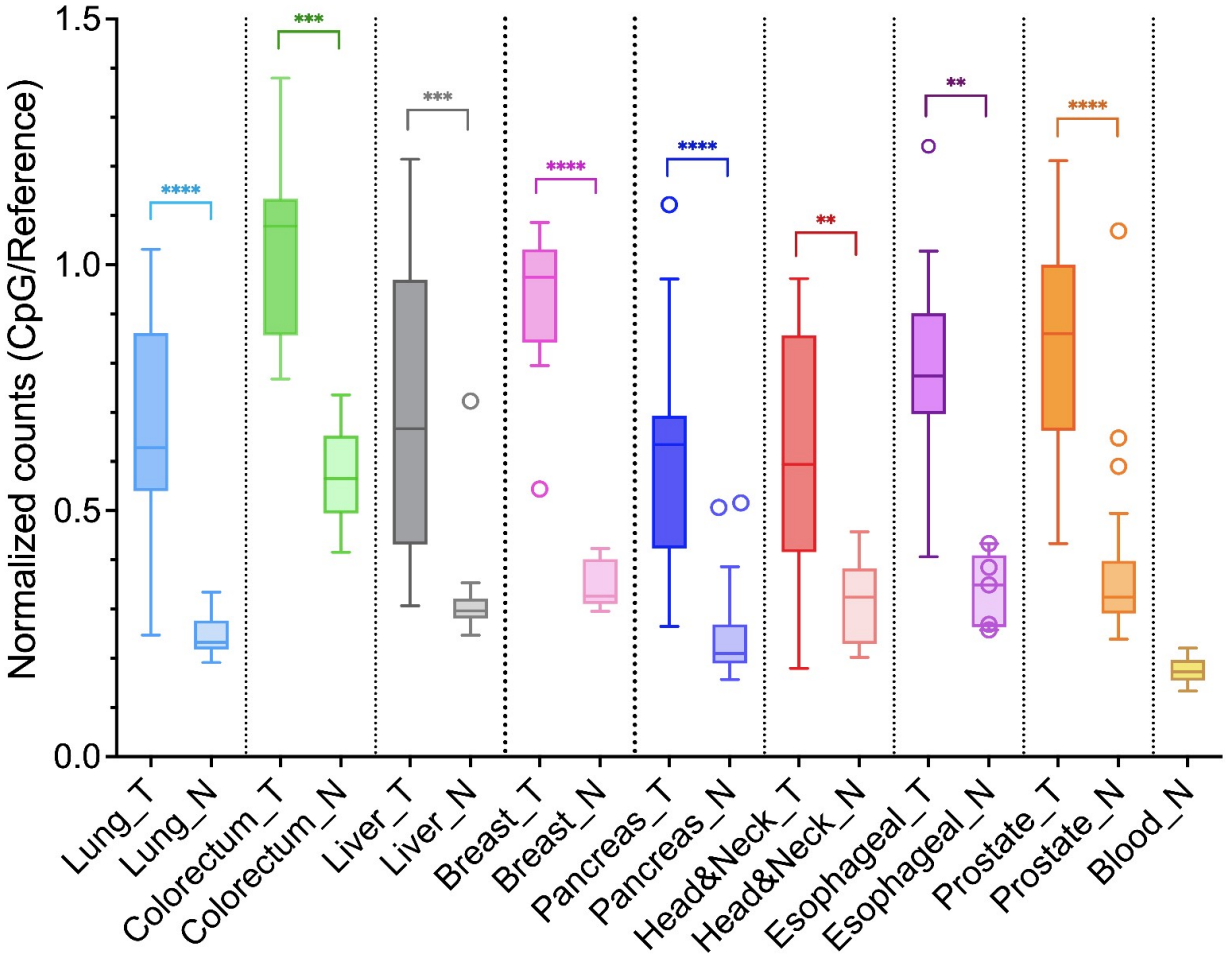


Figure 5 | Sample distribution of normalized counts for all target sites. All nine different tissue types are displayed in distinct colors in the Tukey boxplot. For every tissue type, the sample distribution of the sum of the normalized counts for tumor (T) and normal adjacent tissue (N) are shown. This value is the ratio of the sum of the counts of the CpG smMIPs and the sum of the counts of the reference smMIPs. Significance is indicated with asterisks: * = p-value ⩽ 0.05, ** = p-value ⩽ 0.01, *** = p-value ⩽ 0.001 and **** = p-value ⩽ 0.0001. Mann-Whitney U test was performed using GraphPad Prism.

Selection of the most efficient and discriminating smMIPs

To determine the final biomarker panel for the classifier model, the most efficient and discriminating smMIPs were selected. Using the final dataset with counts of 2,312 CpG targeting smMIPs for 259 samples, a single smMIP linear discriminant analysis (LDA) model was constructed using 5-fold cross validation and the mean cross validated AUC (cvAUC) was calculated for each smMIP. ROC curves for three selected smMIPs are shown in Fig. 6. The distribution of cvAUC is shown in Suppl. Fig. 4. We used a cvAUC cutoff of 0.8 for selecting the best differentiating smMIPs, and a cutoff of 1,000 cumulative counts per smMIP in all undigested samples as a measure for smMIP efficiency. This resulted in a set of 511 CpG smMIPs. Additionally, for CpG sites targeted by multiple smMIPs (i.e. double-tiled), the best performing one was selected. The 511 remaining CpG smMIPs targeted 358 CpG sites. Of these sites, 153 were targeted double-tiled. The difference in cvAUC between smMIPs targeting the same CpG site was less than 0.05 for 84.3% of the multi-targeted CpG sites. The correlation coefficient of these cvAUC values is r=0.668 (Suppl. Fig. 5). Finally, 358 single-tiled CpG smMIPs remained for further analyses. For the reference smMIPs, only the efficient smMIPs were selected, with the same count cutoff of 1,000 cumulative counts in all undigested samples. As a result, 529 reference smMIPs were selected.

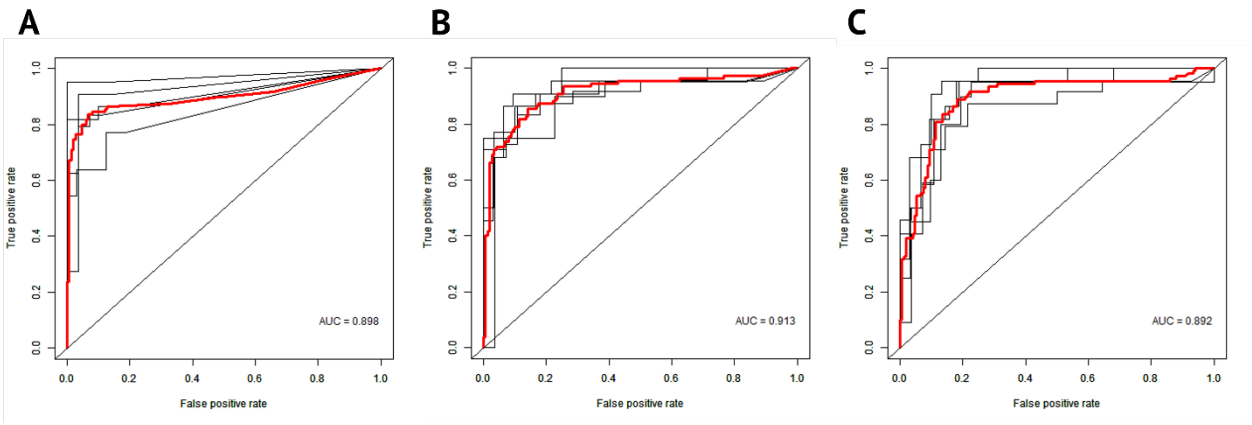


Figure 6 | ROC curves of the three selected target sites for the IMPRESS assay. For each model, the 5-fold cross validation ROC curves (black) and the mean ROC curve (red) are plotted. AUC values are shown for each ROC curve. False positive rate = 1-Specificity, True positive rate = Sensitivity

Classifier model

The remaining 358 CpG smMIP models were then combined into a single model by first selecting the cutoff for every single model for which the sum of false positives and false negatives was the lowest. Then all predictions were combined, and the cutoff was selected based on the highest overall accuracy, which was achieved when 114 single smMIP models agreed on a tumor classification. This final model has a sensitivity of 0.95, a specificity of 0.91, and an accuracy of 0.92 (Table 2).

In addition, we investigated the results per cancer type (Table 2). Accuracy per cancer type ranged from 0.88 to 1, with the exception of colorectal cancer, which performed significantly worse than all other types (accuracy of 0.47). Sensitivity was very high overall, with only 6 false negatives among liver, pancreas and head and neck tumors. False positives are attributed to five tumor types, with colorectal tumors among the highest (specificity 0.10), which skews the overall specificity. However, when exclusively investigating colorectal samples, the cutoff can be adjusted to 282 single smMIP models to obtain a classification accuracy of 1. Interestingly, healthy blood samples never showed up as false positives.

Table 2 | Metrics of our classifier model.

	Multi-cancer	***Lung***	***Colorectal***	***Liver***	***Breast***	***Pancreas***	***Head and neck***	***Esophagus***	***Prostate***	***Blood***
True positives	104	21	7	11	10	19	11	10	15	-
True negatives	135	22	1	10	8	22	5	5	24	35
False positives	14	0	9	1	1	2	0	0	1	0
False negatives	6	1	0	2	0	1	2	0	0	-
Sensitivity	0.945	0.955	1.000	0.846	1.000	0.950	0.846	1.000	1.000	-
Specificity	0.906	1.000	0.100	0.909	0.889	0.917	1.000	1.000	0.960	1.000
Accuracy	0.923	0.977	0.471	0.875	0.947	0.932	0.889	1.000	0.975	-
Balanced accuracy	0.926	0.977	0.550	0.878	0.945	0.934	0.923	1.000	0.980	-

The multi-cancer column combines the results of all tumor types.

Potential for liquid biopsies

To test whether the technique holds potential for use as a multiplex tool for methylation detection in liquid biopsy samples, several characterization experiments were performed. The results are described below and are shown in Fig. 7.

Determination of the amount of input DNA

To test the possibility of lowering the DNA input amount of the IMPRESS technique, we tested different reaction conditions for the MSRE digestion and the smMIP sequencing. The results indicate the feasibility of lowering input amount to 5ng and 10ng DNA for the digest and the smMIP sequencing, respectively (lowest amounts tested) (Suppl. Methods and Suppl. Fig. 6).

Cell free DNA

To further explore the lower limits of input DNA and the feasibility of the technique to study liquid biopsies, four cell-free DNA (cfDNA) samples were tested with 5ng of input. Besides MSRE digested samples, undigested samples were also included as positive controls. The samples were captured by smMIPs and sequenced

according to our protocol. The sequencing results were analyzed using our in-house developed pipeline and read counts were obtained for the 2,331 CpG smMIPs and 600 reference smMIPs. To normalize for the DNA input, the sum of CpG smMIP counts was divided by the sum of reference smMIP counts for each sample. Results showed that digested samples had an average normalized count of 0.24 while the undigested samples had an average normalized count of 2.24 (Fig. 7a). This indicates that the samples were effectively digested by the MSREs as well as efficiently captured by the smMIPs and sequenced.

To mimic the presence of circulating tumor DNA (ctDNA) in cfDNA, DNA from three tumor cell lines was sheared into fragments of 150-500bp and spiked into cfDNA samples in different percentages between 0% and 100%. The calculated percentages based on the normalized counts (see Suppl. methods) were closely correlated to their expected value, with a correlation coefficient r of 0.97, 0.99 and 0.98 for the three different cell lines (Fig. 7b). This indicates the applicability of the technique for the quantification of methylation. The calculated percentage of 20% spiked-in DNA (lowest percentage tested) ranged from 21% to 30%, while those of 0% spiked-in DNA (only cfDNA) ranged from 6% to 13%. This means samples with 20% spike-in of cell line DNA had on average a threefold higher percentage of normalized counts than cfDNA samples without spike-in.

Limit of methylation detection

To determine the limit of methylation detection, we spiked human fully methylated DNA into non-methylated DNA in different percentages between 0% and 12.5% in triplicate. All samples were digested, captured by smMIPs and sequenced following our protocol. The samples with a low methylation level only resulted in background signal. We calculated the limit of blank (LOB) and limit of detection (LOD) following the formulas described by Armbruster *et al.* [27] to determine the lowest detectable methylation level. The LOD corresponds to a methylation level of 4.04% (Fig. 7c). For samples with methylation levels above 4.04%, the measurements will exceed the background signal.

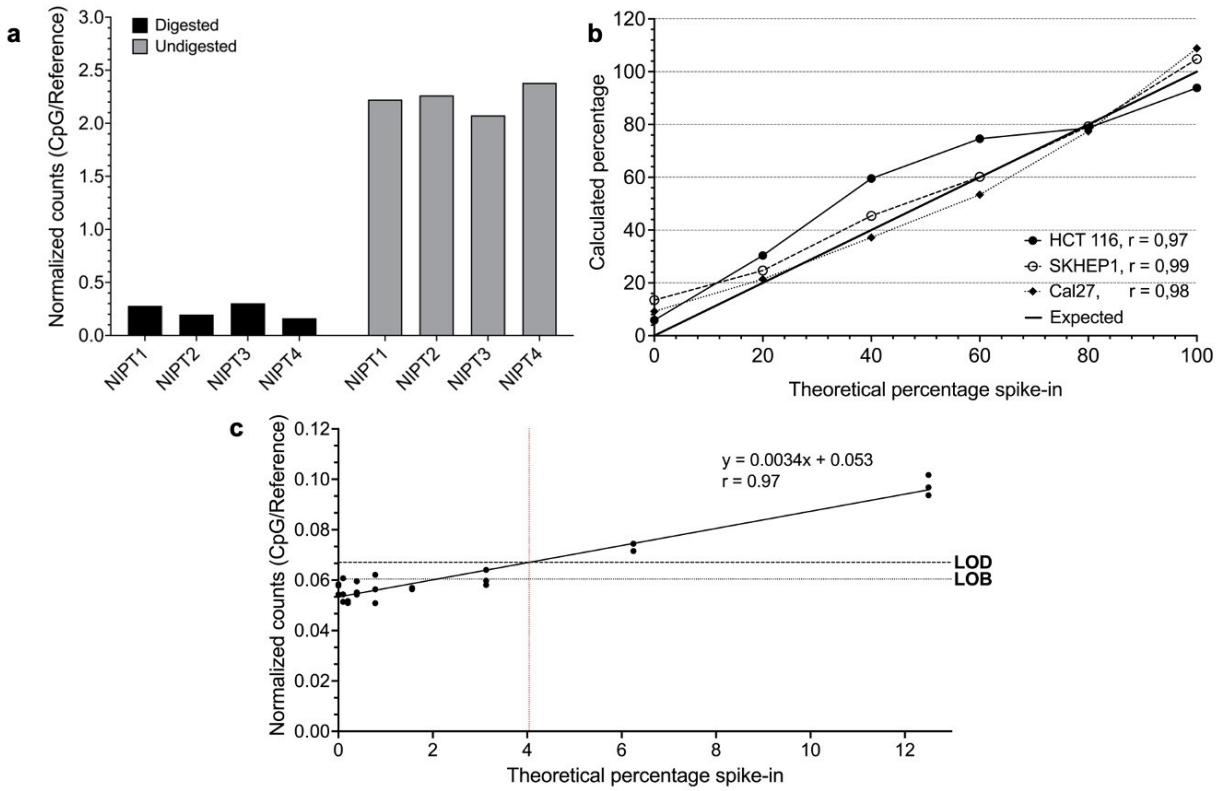


Figure 7 | Potential for liquid biopsies. a) Normalized counts of digested and undigested cfDNA samples. Four cfDNA samples were used to verify that the protocol works in liquid biopsies. On the left side, digested samples are shown. On the right side, undigested samples are shown as positive controls. The DNA input amount is 5ng. b) Calculated percentages based on normalized counts for all target sites of normal cfDNA with spiked-in cell line DNA. Sheared cell line DNA was spiked into normal cfDNA at 0-20-40-60-80-100%. The percentages are calculated by the linear regression through the normalized counts for each sample. The expected value ranges from 0% to 100%, although it is known that normal samples are not 0% methylated in all our targets. The correlation of the calculated and expected percentages is given by r for each cell line. See Suppl. methods for calculations. c) Limit of methylation detection calculated by the normalized counts for all target sites. Fully methylated human DNA was spiked into unmethylated human DNA at 0-0.10-0.20-0.39-0.78-1.56-3.13-6.25-12.5% in triplicate. The linear regression through the normalized counts for each sample has a Pearson correlation coefficient of r=0.97. The limit of blank (LOB) is 0.060 and the limit of detection (LOD) is 0.067. The methylation percentage corresponding to the LOD is 4.04%. Linear regression and plotting were performed using GraphPad Prism.

DISCUSSION

In this study, we successfully created a novel methylation detection technique by combining MSRE digestion with smMIPs. Although digestion of DNA with MSREs has been described since 1978 [28] and has been standard practice for the past decades, smMIPs were only described less than ten years ago [16]. To date, the use of smMIPs was limited to the detection of DNA mutations, microsatellite instability, gene amplifications and differential expression (cDNA) [17,25,29]. In this regard, smMIPs have frequently been proposed for use in routine diagnostics for cancer detection in recent years. In concordance with our results, several studies have demonstrated the high sensitivity of smMIP-sequencing, as well as the possibility to use limited amounts of material [17,18,29–32]. smMIPs have also demonstrated their utility in other research fields [33,34]. smMIP panels can be easily adapted towards their intended application, which is extremely useful with the rapid expansion of molecular markers in all fields [17]. Taken together, this places our novel technique as a promising and widely applicable epigenetic tool for the detection and follow-up of many diseases.

We developed a multi-cancer methylation biomarker panel, and we validated this panel by combining it with the IMPRESS technique, resulting in a robust multi-cancer detection assay. With an overall cross-validated accuracy of 0.92, this final model performed very well in classifying samples. The overview of the sample distribution (Fig. 5) shows a spread of tumor samples for most of the tumor types, while normal tissues are grouped closer together. There was no correlation between tumor cell percentage and normalized counts (data not shown). The spread of tumor samples is most noticeable for head and neck, and liver tumors. The former is a heterogeneous group of locations and cell types, which can account for this spread. For liver tumors, there is only one cell type, but there was a lot of variety in the patient group. Clinical records show that some of the samples originated from patients with alcohol abuse and/or patients with hepatitis. This could potentially have affected the methylation levels in the liver.

Most false positive predictions are the result of normal adjacent colorectal tissue samples having a higher methylation for our targets than the other normal tissue groups (Fig. 5). This could be due to field cancerization, which causes (epi)genetic alterations in histologically normal-looking tissue adjacent to cancerous lesions [35]. Although this phenomenon has also been described in other cancer types, this is not clearly seen in our analyses. In Figure 5, colorectal samples have the highest methylation rate, both for normal samples and tumor samples. As a result, perfect separation of tumor and normal samples is observed when taking only colorectal samples into account, while many normal colorectal samples are false positives in the overall model. This suggests that colorectal cancer might not be a suitable addition to a multi-cancer assay utilizing this biomarker panel, despite its potential differentiating ability within colorectal samples.

While the final model does not make a correct prediction for all samples (0.95 sensitivity and 0.91 specificity), for every sample there are single-smMIP models that do. This emphasizes the exceptional performance of our biomarker panel and confirms the presence of significant methylation differences across all samples.

This model was specifically constructed to include many CpG sites because we intend to use this assay in liquid biopsies in the future. There, only a limited amount of ctDNA is available in the cfDNA, especially in early tumor stages. This ctDNA is fragmented and chances are small that the whole tumoral genome is covered in a liquid biopsy sample.

With respect to liquid biopsies, we performed some additional characterization experiments to thoroughly test the amount of input DNA for the IMPRESS technique. For the digestion reaction, the performance of the MSREs remained equal when lowering the input amount to 5ng. We also showed that an input of 10ng can discriminate between tumor cell lines and healthy blood samples equally well as 20ng and 100ng (Suppl. Fig. 6). Moreover, the use of 5ng cfDNA has been successfully tested in our lab (Fig. 7a). This is extremely important in a liquid biopsies context, where often even less than 5ng is available. To mimic cfDNA, we experimented with sheared tumor cell line DNA that was spiked into normal cfDNA samples. The results demonstrated the possibility to discriminate 20% spiked-in tumor DNA from normal cfDNA. The high correlation coefficients indicated the quantification potential of the technique. However, the counts value is always a relative number and not an exact methylation percentage, as the majority of the targeted regions contain several CpG sites. In theory, we only sequence targets in which all the CpG sites are methylated and as such, we measure the count of all fully methylated fragments. Finally, we did a limit of detection experiment. As there was some background signal, we determined the methylation value for which the measurement exceeded the background signal. This limit of methylation detection was 4.04%. A potential explanation for the background signal is incomplete MSRE digestion. We know that a small percentage of the DNA is not properly digested. This is estimated to be 1.3% by the internal control. In addition, commercially purchased 0% methylated DNA was used, from which previous data (not shown) indicate that a small degree of methylation is still present.

In recent years, multi-cancer detection (MCD) has gained more interest. The IMPRESS technique combined with the multi-cancer biomarker panel shows great potential in this field. Most MCD tests are described for liquid biopsies, which is yet to be done for IMPRESS. Nevertheless, preliminary results show that IMPRESS could become an important and novel addition to the liquid biopsy field.

Cohen *et al*. described their CancerSEEK test in 2018. They detected both circulating proteins and mutations in cfDNA of eight different cancer types [36]. Although mutations and proteins have often been the first choice when developing biomarker assays, it has become clear that methylation signatures possess some major advantages compared to the former. Methylation occurs very early in cancer development, possibly before actual neoplastic transformation, which renders it especially interesting for early diagnosis. Given that no prior

knowledge is needed of the tumor molecular profile, methylation biomarkers are more universal than mutation markers. Since methylation-based tests can be used off the shelf, they are much faster and cheaper to use [37] . The utility of methylation signatures was for example demonstrated by Chen *et al*., who published the PanSeer test in 2020. The results obtained with their blood-based test show that cancer is detected in 95% of asymptomatic patients for five cancer types. Although further investigation is needed to confirm the results, they claim that several cancers can already be detected four years prior to diagnosis, based on methylation biomarkers [38], which is very promising for early cancer detection. Others have worked on the use of methylation signatures for screening and early detection as well. The biotechnology company GRAIL finances several clinical trials where the use of their Galleri® test is evaluated. These trials were started based on publications by Liu *et al.* and Klein *et al.,* who first tested and independently validated the performance of targeted methylation analysis of cfDNA for multiple cancer types [37,39] . Both for the PanSeer as well as for the first experiments with the Galleri® test, bisulfite-based technologies were used [37–39]. We believe that the sensitivity in cfDNA could be increased with our novel technique using MSREs, as we avoid bisulfite conversion, which is a harsh chemical treatment of DNA.

The use of enzymes has recently been gaining more attention. A prime example is the recent development of the Enzymatic Methyl sequencing (EM-seq) technology. This EM-seq technology is for example used for targeted methyl-seq in combination with the Twist Human Methylome Panel (Twist Bioscience). In a limited number of publications, EM-seq is described to be superior to WGBS for sensitivity, repeatability and base composition [11,13,14]. However, EM-seq requires a large investment for library preparation[11] . TAPS was also recently described in a few papers [12,40,41]. However, there are no external publications comparing TAPS to other state of the art technologies in literature. Considering smMIPs, Arts *et al.* already described the low cost of smMIP-sequencing, which does not drastically increase with MSRE treatment. Therefore, our novel technique is more cost-effective compared to current bisulfite-free alternatives on the market [25].

The application of the IMPRESS technique lies in targeted biomarker sequencing. This type of targeted sequencing technology is becoming more popular, and our technique can become an important new platform to be used in this area. IMPRESS enables the multiplexing of a considerable number of target sites, extending to thousands, in contrast to methodologies like droplet digital PCR (ddPCR), which offer limited multiplexing capabilities. Additionally, compared to genome-wide techniques, our approach presents a distinct advantage in terms of cost-effectiveness. By selectively analyzing predefined sites of interest, we substantially decrease sequencing costs. The technique is easily implementable in standard equipped laboratories. We use widely available reagents, and the protocols are straightforward, making the technique easily applicable for research groups with access to NGS infrastructure. The hands-on time from DNA extraction to sequencing is approximately 4.5 hours, which is comparable with EM-seq [42]. Moreover, by sequencing the insert fragment of interest, we can detect and correct for nonspecific hybridization. Notably, our technique boasts high

throughput, facilitating the simultaneous analysis of 9x384 samples within a single sequencing run. Depending on the desired coverage per sample, a larger sequencing kit may be required. Furthermore, this technique has the potential for integrating different genetic information into one assay, for example, mutation and CNV analysis. smMIPs could be designed for any target type, taking the MSRE recognition sites into account, and combined into a single assay. As such, one single experiment could provide the information that is now only obtained after several analyses.

There are a few limitations in this study. The most important one is that we have not yet tested liquid biopsies from cancer patients. However, our results show that blood samples register very low normalized counts by the IMPRESS technique for our biomarker panel. In addition, we can efficiently use 5 ng of cfDNA as input, and there is a limit of methylation detection of 4.04%. In the future, more optimization steps will be executed, and liquid biopsy samples from cancer patients will be tested.

Another limitation is that only 39% of the CpG sites of the methylome are located in recognition sites of the enzymes used in our assay[9]. This could easily be solved by using other or additional restriction enzymes, through which a large majority of CpG sites in the genome can be made available for analysis.

Furthermore, we could not include tissue from fully healthy persons, instead, we used normal tissue collected at a distance from the tumor. Clinical records do not show the exact distance. In the literature, it has been described that tissue samples adjacent to a tumor might have molecular alterations (e.g. field cancerization) but look microscopically like normal tissue [43,44]. This might result in a more difficult discrimination between normal and cancer tissue by our classifier model.

Lastly, we did not yet include cancer type-specific smMIPs for the detection of tissue-of-origin. Classification of cancers of unknown primary is an important aspect of cancer diagnostics and is increasingly described in the literature. It can already be determined based on *in silico* analyses [45] and therefore, including TOO determination in our assay is an important step for the future. Analyses have been performed in our research group to determine tissue-specific methylation patterns, for which smMIPs will be designed and included in the future. This will not cause any problems as we have already demonstrated the very high multiplex capacity of our technique.

In conclusion, we developed a novel method for sensitive detection of DNA methylation, we developed a multi-cancer methylation biomarker panel, and we combined those two into a multi-cancer detection assay. Our characterization experiments already demonstrate the application of this technique and the biomarker panel for low amounts of fragmented DNA. The combination of multiple markers covered by the smMIPs allows for the high sensitivity that is essential in liquid biopsies and early cancer detection applications.

DECLARATIONS

Acknowledgements

We would like to thank all patients and the UZA biobank for their contribution to this study. In addition, we would like to thank Bronwen Martin to read our manuscript and improve the English grammar and academic vocabulary.

Authors' contributions

JV: Conceptualization of the study, methodology, interpreting results, writing original draft, reviewing and editing of the final draft, **IN**: Conceptualization of the study, methodology, interpreting results, writing original draft, reviewing and editing of the final draft, **TV**: Conceptualization of the study, methodology, interpreting results, writing original draft, reviewing and editing of the final draft, **JI**: Methodology (CpG site selection), reviewing and editing of the final draft, **ASu**: Conceptualization of the study, interpreting results, reviewing and editing of the final draft, **DP**: Methodology (histopathological analysis), reviewing and editing of the final draft, **ASc**: Methodology (processing of the samples), **AH**: Methodology (design of the smMIPs), **EF**: Methodology (statistics), reviewing and editing the final draft, **MP**: Conceptualization of the study, reviewing and editing the final draft, **GVC**: Conceptualization of the study, interpreting results, reviewing and editing of the final draft, **KODB**: Conceptualization of the study, interpreting results, reviewing and editing the final draft.

All authors have read and agreed to the published version of the manuscript.

Ethics approval and consent to participate

This study was conducted in accordance with Good Clinical Practice guidelines and the Declaration of Helsinki. Fresh frozen tissue samples used in this study were previously collected in the Biobank of the Antwerp University Hospital and retrospectively used in this study. According to Article 20 of the Belgian Law on the procurement and use of human corporal material intended for human application or scientific research of 19 December 2008, patients provide consent for the use of their bodily material in research when consenting to an invasive procedure. As such, no additional consent was needed for the use of these retrospective samples. For prospectively collected blood samples, informed consent was given by each subject. The study protocol and any modifications thereof were approved by the UZA ethical committee (Ref. N°20/02/056 and Ref. N°41/14/426) before experimental analyses were performed.

Consent for publication

Not applicable.

Data availability

The data that support the findings of this study are available on request from the European Genome-Phenome Archive (EGAS00001007559).

Competing interests

The authors declare that they have no competing interests.

Funding

Research in the lab of Prof. Guy Van Camp is supported by grants awarded by the University of Antwerp (IOF/SBO 43782, BOF/Methusalem grant 40790, BOF/TOP 39705). Joe Ibrahim and Isabelle Neefs are supported by a strategic basic PhD fellowship of the Research Foundation Flanders (FWO; 11B5220N and IS67523N).

REFERENCES

1. Globocan I. Cancer Today. Data visualisation tools for exploring the global cancer burden in 2020. 2020.

2. WHO. Cancer - Screening and early detection. 2010.

3. National Cancer Institute. Stage at Diagnosis. 2022.

4. Kulis M, Esteller M. Chapter 2 - DNA Methylation and Cancer. Epigenetics and Cancer, Part A. 2010;70:27–56.

5. Constâncio V, Nunes SP, Moreira-Barbosa C, Freitas R, Oliveira J, Pousa I, et al. Early detection of the major male cancer types in blood-based liquid biopsies using a DNA methylation panel. Clin Epigenetics. 2019;11.

6. Ibrahim J, Op de Beeck K, Fransen E, Peeters M, Van Camp G. Genome-wide DNA methylation profiling and identification of potential pan-cancer and tumor-specific biomarkers. Mol Oncol. 2022;16:2432–47.

7. Martisova A, Holcakova J, Izadi N, Sebuyoya R, Hrstka R, Bartosik M. Dna methylation in solid tumors: Functions and methods of detection. Int J Mol Sci. MDPI; 2021.

8. Liu Z, Wang Z, Jia E, Ouyang T, Pan M, Lu J, et al. Analysis of genome-wide in cell free DNA methylation: progress and prospect. Analyst [Internet]. 2019/08/23. 2019;144:5912–22. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31436778

9. Beikircher G, Pulverer W, Hofner M, Noehammer C, Weinhaeusel A. Multiplexed and sensitive DNA methylation testing using methylation-sensitive restriction enzymes "MSRE-qPCR." Methods in Molecular Biology. Humana Press Inc.; 2018. p. 407–24.

10. New England Biolabs (NEB). MspJI.

11. Sun Z, Vaisvila R, Hussong LM, Yan B, Baum C, Saleh L, et al. Nondestructive enzymatic deamination enables single-molecule long-read amplicon sequencing for the determination of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Genome Res. 2021;31:291–300.

12. Liu Y, Siejka-Zielińska P, Velikova G, Bi Y, Yuan F, Tomkova M, et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. Nat Biotechnol. 2019;37:424–9.

13. Han Y, Zheleznyakova GY, Marincevic-Zuniga Y, Kakhki MP, Raine A, Needhamsen M, et al. Comparison of EM-seq and PBAT methylome library methods for low-input DNA. Epigenetics. 2021;

14. Vaisvila R, Chaithanya Ponnaluri VK, Sun Z, Langhorst BW, Saleh L, Guan S, et al. EM-seq: Detection of DNA Methylation at Single Base Resolution from Picograms of DNA. Available from: https://doi.org/10.1101/2019.12.20.884692

15. Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014. p. 418–26.

16. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res. 2013;23:843–54.

17. Steeghs EMP, Kroeze LI, Tops BBJ, Van Kempen LC, Ter Elst A, Kastner-Van Raaij AWM, et al. Comprehensive routine diagnostic screening to identify predictive mutations, gene amplifications, and microsatellite instability in FFPE tumor material. BMC Cancer. 2020;20.

18. Bekers EM, Eijkelenboom A, Rombout P, Van Zwam P, Mol S, Ruijter E, et al. Identification of novel GNAS mutations in intramuscular myxoma using next-generation sequencing with single-molecule tagged molecular inversion probes. Diagn Pathol. 2019;14.

19. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics [Internet]. 2012;28:2520–2. Available from: https://doi.org/10.1093/bioinformatics/bts480

20. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013;00:1–3. Available from: http://arxiv.org/abs/1303.3997

21. Broad Institute. Picard Toolkit [Internet]. 2019. Available from: https://broadinstitute.github.io/picard/

22. Boyle EA, O'Roak BJ, Martin BK, Kumar A, Shendure J. MIPgen: Optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics. 2014;30:2670–2.

23. Venables WN, Ripley BD. Modern Applied Statistics with S [Internet]. 4th ed. New York: Springer; 2002. Available from: http://www.stats.ox.ac.uk/pub/MASS4

24. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics [Internet]. 2005;21:3940–1. Available from: https://doi.org/10.1093/bioinformatics/bti623

25. Arts P, Van Der Raadt J, Van Gestel SHC, Steehouwer M, Shendure J, Hoischen A, et al. Quantification of differential gene expression by multiplexed targeted resequencing of cDNA. Nat Commun. 2017;8.

26. Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, Yau P, et al. Microarray-based DNA methylation profiling: Technology and applications. Nucleic Acids Res. 2006;34:528–42.

27. Armbruster DA, Pry T. Limit of Blank, Limit of Detection and Limit of Quantitation. Clin Biochem Rev. 2008.

28. Harrison A, Parle-McDermott A. DNA methylation: A timeline of methods and applications. Front Genet. 2011.

29. Wu D, Waalkes A, Penewit K, Salipante SJ. Ultrasensitive detection of chimerism by single-Molecule molecular inversion probe capture and high-Throughput sequencing of copy number deletion polymorphisms. Clin Chem. 2018;64:938–49.

30. Weren RDA, Mensenkamp AR, Simons M, Eijkelenboom A, Sie AS, Ouchene H, et al. Novel BRCA1 and BRCA2 Tumor Test as Basis for Treatment Decisions and Referral for Genetic Counselling of Patients with Ovarian Carcinomas. Hum Mutat. 2017;38:226–35.

31. Neveling K, Mensenkamp AR, Derks R, Kwint M, Ouchene H, Steehouwer M, et al. BRCA testing by single-molecule molecular inversion probes. Clin Chem. 2017;63:503–12.

32. Verkouteren BJA, Wakkee M, van Geel M, van Doorn R, Winnepenninckx VJ, Korpershoek E, et al. Molecular testing in metastatic basal cell carcinoma. J Am Acad Dermatol. 2021;85:1135–42.

33. Acuna-Hidalgo R, Sengul H, Steehouwer M, van de Vorst M, Vermeulen SH, Kiemeney LALM, et al. Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. Am J Hum Genet. 2017;101:50–64.

34. Šestáková Š, Šálek C, Remešová H. DNA Methylation Validation Methods: A Coherent Review with Practical Comparison. Biol Proced Online. BioMed Central Ltd.; 2019.

35. Croes L, op de Beeck K, Pauwels P, Berghe W vanden, Peeters M, Fransen E, et al. DFNA5 promoter methylation a marker for breast tumorigenesis [Internet]. Oncotarget. 2017. Available from: www.impactjournals.com/oncotarget/

36. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test [Internet]. Available from: https://www.science.org

37. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden M V., Liu MC, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Annals of Oncology. 2020;31:745–59.

38. Chen X, Gole J, Gore A, He Q, Lu M, Min J, et al. Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. Nat Commun. 2020;11.

39. Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Annals of Oncology. 2021;32:1167–77.

40. Cheng J, Siejka-Zielińska P, Liu Y, Chandran A, Kriaucionis S, Song CX. Endonuclease enrichment TAPS for cost-effective genome-wide base-resolution DNA methylation detection. Nucleic Acids Res. 2021;49:E76.

41. Liu Y, Hu Z, Cheng J, Siejka-Zielińska P, Chen J, Inoue M, et al. Subtraction-free and bisulfite-free specific sequencing of 5-methylcytosine and its oxidized derivatives at base resolution. Nat Commun. 2021;12.

42. Morrison J, Koeman JM, Johnson BK, Foy KK, Beddows I, Zhou W, et al. Evaluation of whole-genome DNA methylation sequencing library preparation protocols. Epigenetics Chromatin. 2021;14.

43. Hawthorn L, Lan L, Mojica W. Evidence for field effect cancerization in colorectal cancer. Genomics. 2014;103:211–21.

44. Cherkezyan L, Stypula-Cyrus Y, Subramanian H, White C, Dela Cruz M, Wali RK, et al. Nanoscale changes in chromatin organization represent the initial steps of tumorigenesis: a transmission electron microscopy study. BMC Cancer. 2014;14:189.

45. Sun K, Wang J, Wang H, Sun H. GeneCT: A generalizable cancerous status and tissue origin classifier for pan-cancer biopsies. Bioinformatics. 2018;34:4129–30.

SUPPLEMENTARY MATERIAL

Supplementary methods: IMPRESS protocol

MSRE digest

The first step was a combined digestion of the DNA with four MSREs (HpaII, HpyCH4IV, AciI and HinP1I). The MSREs digest unmethylated DNA at their recognition sites (C^CGG, A^CGT, C^CGC and G^CGC, respectively). Methylated CpG sites block the restriction enzymes which results in unaffected CpG regions.

In a total digestion volume of 10µL, an input of 50ng DNA of each sample was diluted in 8.1µL H_2O (on ice). Per sample, a mastermix was added containing 1µL CutSmart Buffer (New England Biolabs), 0.5µL EcoRI digested lambda DNA (0.75pg/µL) (Thermo Fisher Scientific), and 0.1µL of each restriction enzyme HpaII, HpyCH4IV, AciI and HinP1I (10U/µL) (New England Biolabs). Lambda DNA was digested by FastDigest EcoRI (Thermo Fisher Scientific) according to the manufacturer's protocol to remove concatemers and spiked-in as an internal control. The reaction was incubated for 16 hours at 37°C, followed by a heat inactivation step for 30 min at 65°C and stored at 4°C until continuation. To enable downstream calculation of the efficiency of our smMIP-pool and the percentage of non-digestion, undigested samples were also included. A modified version of the protocol was executed for all these samples, where the 4 MSREs were replaced by H_2O. All undigested doubles were incubated as described above.

Pooling and phosphorylation

One smMIP pool was made by adding the 2953 smMIPs equimolarly (100µM per smMIP). smMIPs were produced by Integrated DNA technologies (IDT, Coralville, USA). Because the DNA ligase in the capture reaction needs a 5' phosphorylated end, the smMIPs were first phosphorylated. The pool was phosphorylated in the following 35µL reaction composition: 29.53µL smMIP pool (100µM), 1.18µL T4 Polynucleotide Kinase (10U/µl) (New England Biolabs), 3.5µL 10X T4 DNA ligase buffer with 10mM ATP (New England Biolabs), and 0.79µL H_2O. This reaction was incubated for 45 min at 37°C, followed by an inactivation step at 65°C for 20 min, and storage at -20°C until further use. The final concentration of each smMIP in this pool was 0.0286µM (100µM x 0.01µL / 35µL).

Capture

After MSRE digest, methylated CpG regions remained intact and were captured by the smMIPs through hybridization of the smMIP binding arms to the complementary genomic sequence. Elongation and ligation of the smMIP created a circular DNA fragment. In all capture reactions, some smMIPs were ligated without a 50nt insert. These so-called empty smMIPs were filtered out later.

For the capture reaction, a ratio of 800 smMIP copies per DNA target copy was chosen based on previously published research and optimization experiments. The volume of the phosphorylated smMIP pool to add was calculated as described by Arts *et al*. (19) and resulted in a volume of 0.000766µL. The phosphorylated pool was diluted using EB buffer (Qiagen, Hilden, Germany) for practical reasons.

To each 10µL digestion reaction, a 15µL capture mastermix was added (on ice), containing 7.66µL of the 1/10,000 smMIP pool dilution, 2.5µL 10X Ampligase DNA Ligase buffer (Lucigen), 0.032µL dNTPs (0.25nM) (Invitrogen), 0.32 Hemo Klentaq (10U/µL) (New England Biolabs), 0.01µL Ampligase DNA Ligase (100U/µL) (Lucigen), and 4.48µL H_2O. This reaction was incubated for 10 min at 95°C and for 21.5 hours at 60°C. After incubation, the reaction was cooled down on ice for a few minutes and the exonuclease treatment was performed immediately.

Exonuclease treatment

Subsequently, an exonuclease treatment was performed. In this manner, all linear fragments such as unbound smMIPs or original DNA strands were degraded, and only circular fragments stayed intact. Removal of these linear fragments was important to specifically amplify the ligated smMIPs in the subsequent PCR reaction.

For this reaction, 2µL of EXO treatment mastermix was added to each capture reaction (on ice), containing 0.5µL EXO I (20U/µL) (New England Biolabs), 0.5µL EXO III (100U/µL) (New England Biolabs), 0.2µL 10X Ampligase DNA Ligase buffer (Lucigen), and 0.8µL H_2O. Incubation was performed for 45 min at 37°C, followed by a heat inactivation step of 2 min at 95°C. The reaction was stored at 4°C until further use.

qPCR and PCR amplification

Before PCR amplification was initiated, a qPCR experiment was executed to determine the number of PCR cycles needed for a specific smMIP pool. For this one-time qPCR, 20µL of a mastermix containing 12.5µL iProof (Bio-Rad), 0.125µL forward primer (100µM), 0.125µL reverse primer (100µM), 0.125µL SYBR Green (Invitrogen) and 7.125µL H_2O was added to 5µL of each exonuclease treated sample. The following PCR protocol was used: 30s at 98°C, 35 cycles of [10s at 98°C, 30s at 60°C, 30s at 72°C], and 2 min at 72°C.

Thereafter, the circular fragments were amplified by PCR using primers complementary to the universal PCR primer sites in the common smMIP backbone. The primers contained a unique barcode for each sample so many samples could be combined into a single library. The primers also contain a P5 or P7 oligo to allow binding to the Illumina flow cell (suppl figure 1).

The protocol was optimized by executing multiple PCRs reactions. Performing four replicate PCRs allows the effective use of all input DNA. As such, the number of PCR cycles was reduced, leading to less PCR duplicates and increasing the percentage of unique sequencing reads.

For the PCR, a mastermix was made per 192 samples. Each mastermix contained a forward primer identical per 192 samples. In total, 4 different forward primers were combined with 192 unique reverse primers. Firstly, a total of 18.75µL mastermix containing 12.5µL 2X iProof (Bio-Rad), 0.125µL forward Primer (100µM) and 6.125µL H_2O was added to 6.5µL of the exonuclease treated samples. Secondly, 1.25µL of a unique reverse primer (10µM) was added to each sample. The PCR reaction was repeated four times, to use all exonuclease treated product. As such, less PCR cycles were needed to still obtain enough input DNA for sequencing, resulting in less PCR duplicates. The following PCR protocol was used: 30s at 98°C, 19 cycles of [10s at 98°C, 30s at 60°C, 30s at 72°C], and 2 min at 72°C. The products were stored at 4°C until further use.

Forward primer:

AATGATACGGCGACCACCGAGATCTACACNNNNNNNNATACGAGATCCGTAATCGGGAAGCTGAAG

Reverse primer:

CAAGCAGAAGACGGCATACGAGATNNNNNNNNACACGCACGATCCGACGGTAGTGT

Pooling, bead purification and sequencing

Finally, all samples were pooled and purified. The bead purification aims to remove superfluous enzymes, buffers, dNTPs and unnecessary fragments. Low amounts of DNA at the start of the protocol, result in increased amounts of empty smMIPs formed during the capture reaction. These empty smMIPs are formed by ligation without insert elongation. To make optimal use of the sequencing capacity, the empty smMIP fragments are removed before sequencing.

The total of 26.5µL PCR product from each sample was pooled per 24 samples. AMPure XP beads (Beckman Coulter) were added in 1:1 volume ratio, and the reaction was washed twice with 70% ethanol. Subsequently, the DNA of 96 samples was eluted in the same 100µL of TE buffer (0.01M Tris-HCl, 0.001M EDTA). A second purification step was executed to concentrate the pool and to lower the amount of smMIPs that did not capture any target (empty smMIPs), increasing sequencing capacity for correctly captured smMIPs. This is especially important for low DNA input amounts, as empty smMIPs are more prominent here. In the second purification step, 25µL of four 1x-purified pools with the same samples from four replicate PCRs were pooled, and beads were added in a 1:1 volume ratio. This reaction was washed twice with 70% ethanol and eluted in 40µL TE buffer. Optionally, more purification steps could be executed in case there is still an excess in amount of empty smMIPs. Finally, all individual pools were combined into the final NGS library.

Subsequently, concentrations and fragment lengths of the purified pools were determined by a High Sensitivity D1000 Kit on the TapeStation (Agilent). The concentrations of the region between 100bp and 1000bp were used. The final NGS library was sequenced on a High Output Kit on an Illumina NextSeq 550 system. 1.2pM of

the NGS library was loaded and 5% of PhiX DNA was spiked into the library. Libraries were paired-end sequenced (2x75nt) using custom MIPs primers (suppl figure 1). For the validation experiment to evaluate the repeatability, the final library was sequenced using a Nano v2 kit on the Illumina MiSeq System.

Supplementary methods: Determination of the amount of input DNA

For digestion

The standard protocol for an MSRE digest requires 1µg of input DNA in a final reaction volume of 50µL with 10 units of each enzyme. To test the possibility to lower the input amount, we tested different reaction conditions (suppl Figure 6a). The digested samples were amplified by qPCR with primer pairs hybridizing around the MSRE recognition sites. The results show that all tested reaction conditions performed similarly, as the Ct values ranged from 20.4 to 25.1 for primer pair 1, from 22.8 to 25.1 for primer pair 2 and from 28.0 to 30.8 for primer pair 3, while the undigested samples reached the threshold in the range of 11.0 to 12.7 cycles. The results also indicate the feasibility of lowering input DNA to 5ng.

For capturing and sequencing

To test the possibility of the smMIP capturing and sequencing with varying quantities of input DNA, the performance of a reaction starting from 10ng and 20ng was compared to the standard 100ng for smMIP-sequencing (15). gDNA from two blood samples and two cell lines was MSRE digested, captured by smMIPs and sequenced following our protocol. The sequencing results were analyzed using our in-house developed pipeline and read counts were obtained for the 2,331 CpG smMIPs and 600 reference smMIPs. To normalize the DNA input, the sum of CpG smMIP counts was divided by the sum of reference smMIP counts for each sample. This value is assumed to be higher in samples methylated for our targets (i.e. tumor samples) and lower in samples unmethylated for our targets (i.e. normal samples). Blood samples had an average of 0.18, 0.19 and 0.19 normalized counts for 100ng, 20ng and 10ng respectively, while cell lines had an average of 1.10, 1.23 and 1.26 counts respectively (suppl Figure 6b). This demonstrates that the results of our technique are consistent between an input amount of 10ng, 20ng and 100ng.

Supplementary methods: Calculated percentage after spike-in

To mimic the presence of circulating tumor DNA (ctDNA) in cfDNA, DNA from three tumor cell lines was sheared into fragments of 150-500bp using the Covaris. They were spiked into cfDNA samples in different percentages (different conditions): 0% (only cfDNA), 20%, 40%, 60%, 80% and 100% (only sheared cell line DNA). A total of 5ng DNA was used for each condition.

These mock ctDNA-cfDNA mixes were used for the IMPRESS protocol. After data quality control and mapping, read counts were obtained for CpG smMIPs and reference smMIPs in each sample. Normalized counts were calculated by dividing the sum of CpG smMIPs by the sum of reference smMIPs in each sample. For each cell line, a linear regression was plotted through the mock ctDNA-cfDNA samples. The following linear regression equations were obtained, with x being the theoretical percentage spike-in and y being the normalized count:

HCT116	y = -0.6297x + 4.3038
SKHEP1	y = -0.44x + 2.9083
Cal27	y = -0.5934 + 3.6367

These functions were used to predict the normalized counts value (y) for the 100% cell line samples (x = 100) corresponding to the linear regression line. This was necessary because in this way, not only the 100% cell line sample was determinative for the normalized counts value, but all the samples used for the linear regression line were.

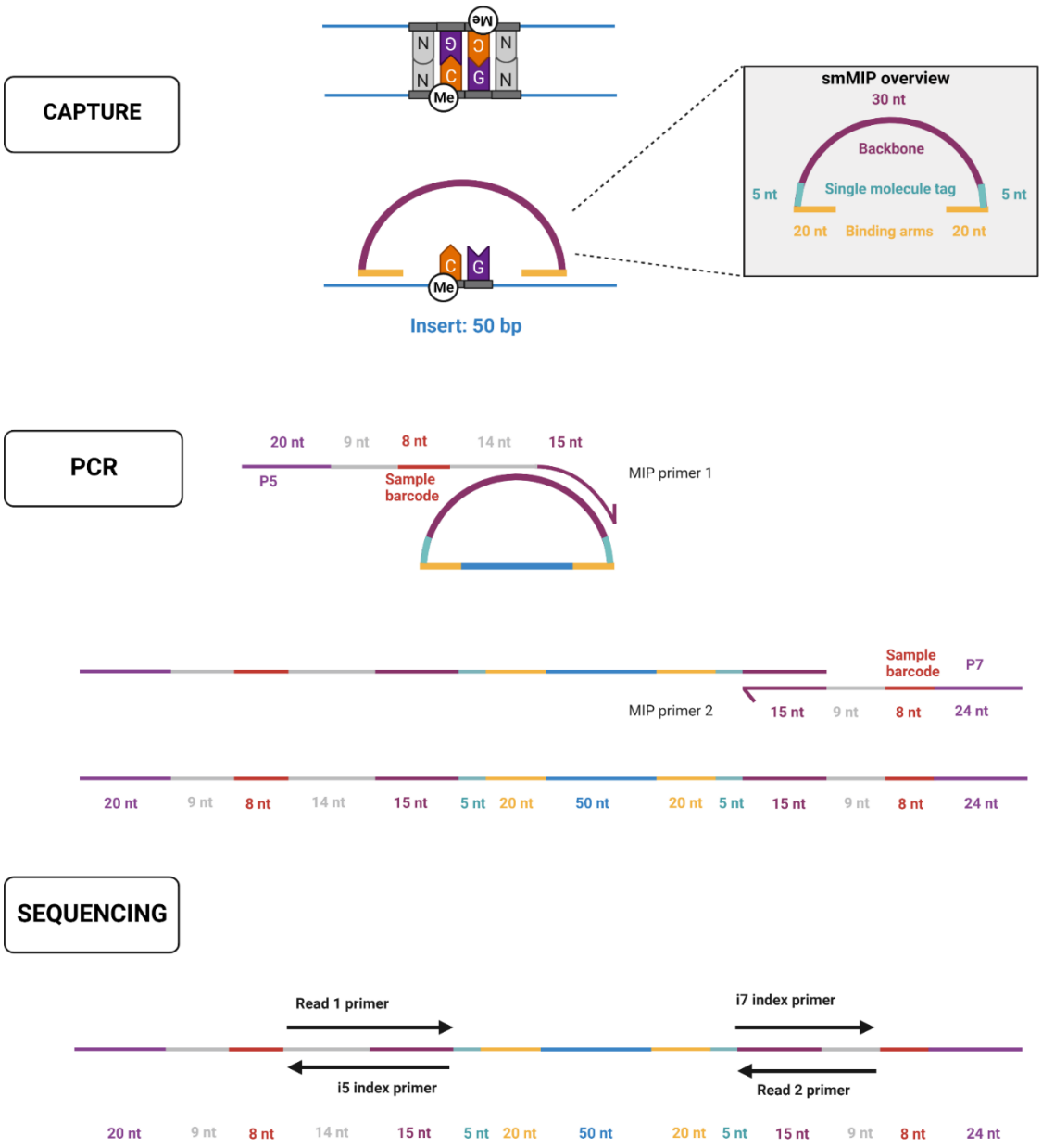
Subsequently, for each sample, the normalized count was divided by the calculated 100% normalized counts value obtained from the linear regression from the corresponding cell line, to calculate the ratio (calculated percentage). This was done because the normalized counts differed per cell line and by calculation of the ratio, comparison of the cell lines was made possible.

Lastly, the Pearson's correlation coefficient of the calculated percentage (i.e. observed spike-in percentage) and the expected percentage (i.e. theoretical spike-in percentage) was calculated.

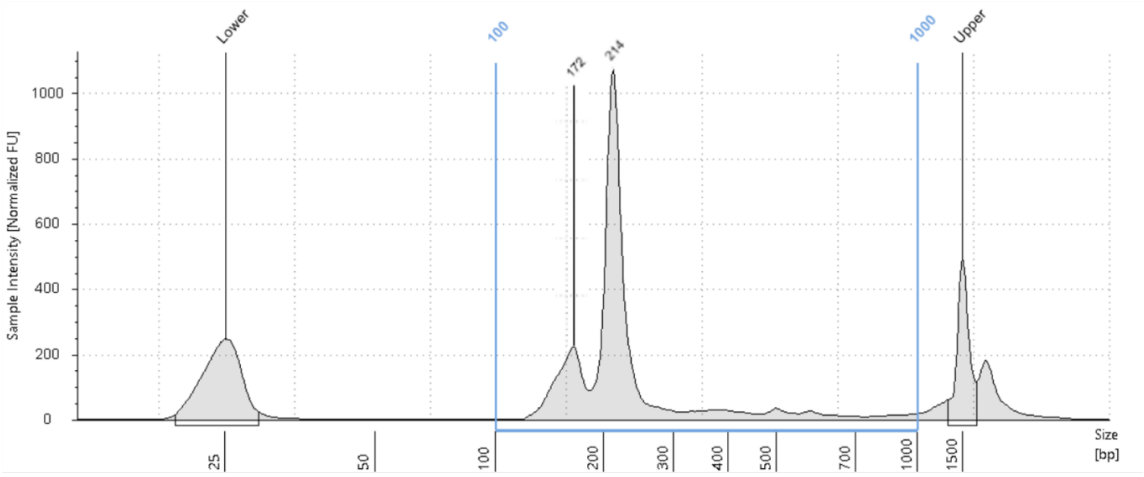
References

1. Boeckx N, Op de Beeck K, Beyens M, Deschoolmeester V, Hermans C, De Clercq P, et al. Mutation and Methylation Analysis of Circulating Tumor DNA Can Be Used for Follow-up of Metastatic Colorectal Cancer Patients. Clin Colorectal Cancer. Elsevier Inc.; 2018;17:e369–79.

Supplementary Figures

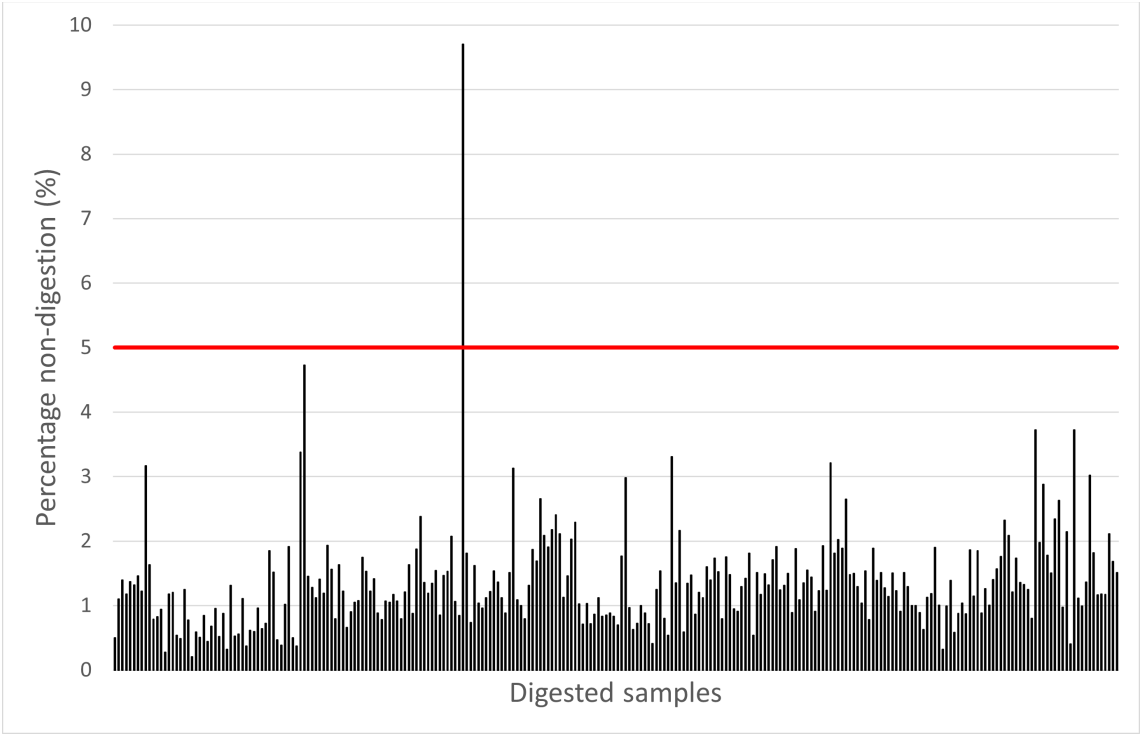


Suppl. Figure 1 | Molecular overview of IMPRESS. Numbers represent the nucleotide length of each fragment. Each smMIP contains a common smMIP backbone of 30nt (pink), single molecule tags of 5nt (green), and two binding arms of circa 20nt (yellow). During capture, the smMIP hybridizes to the target site, in order to have an insert length of 50nt. During PCR, MIP primers bind to the common smMIP backbone (pink). These primers additionally contain a sample barcode (red), a P5 or P7 oligo to bind the flowcell (purple) and supporting fragments (grey). During sequencing, the read 1 and 2 primers are used for the sequencing of the single molecule tag (green), the binding arm (yellow) and the insert fragment (blue), 75nt in each direction. The index primers are used for sequencing of the sample barcodes (red). Created with BioRender.com

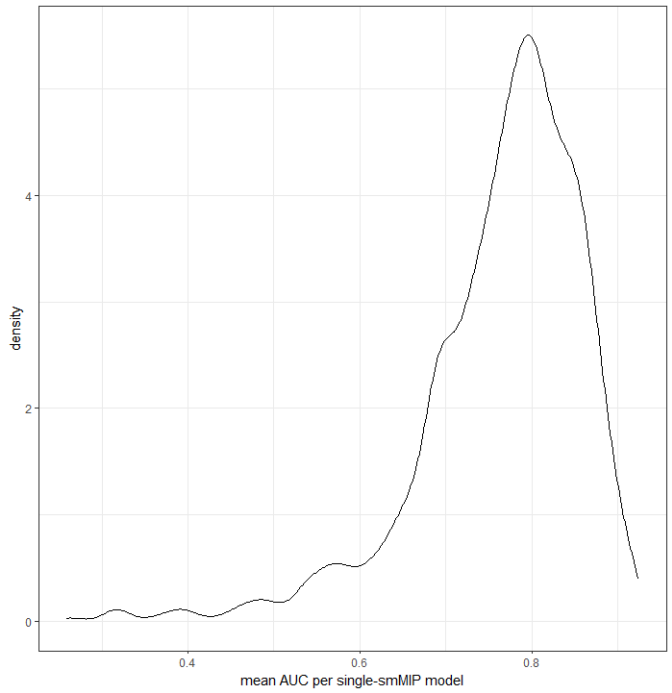


From [bp]	To [bp]	Average Size [bp]	Conc. [pg/µl]	Region Molarity [pmol/l]	% of Total	Region Comment	Color
100	1000	251	1260	8890	85.04		

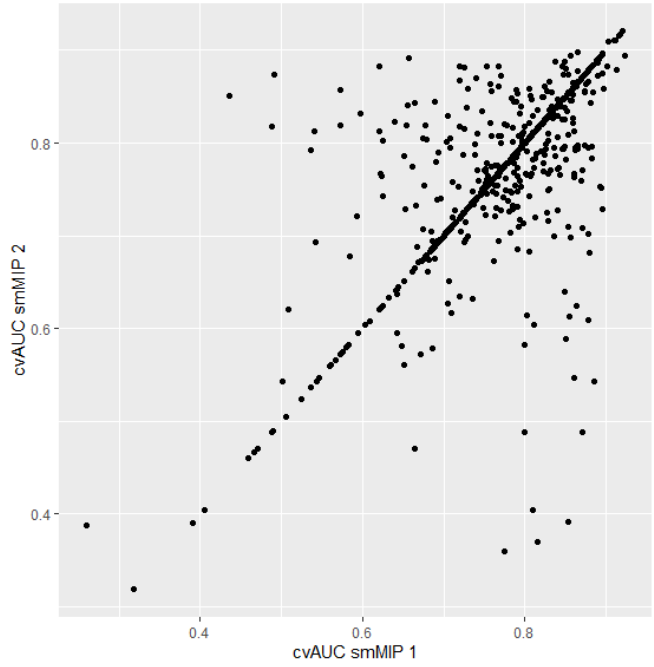
Suppl. Figure 2 | Capillary electrophoresis analysis of the final NGS library using the Tapestation (Agilent). The target peak is seen around 214bp (length of smMIP with insert = 222nt). 50bp to the left, the empty smMIPs are shown. For sequencing, the concentration of the pool is estimated based on the range from 100bp to 1000bp to avoid overclustering.



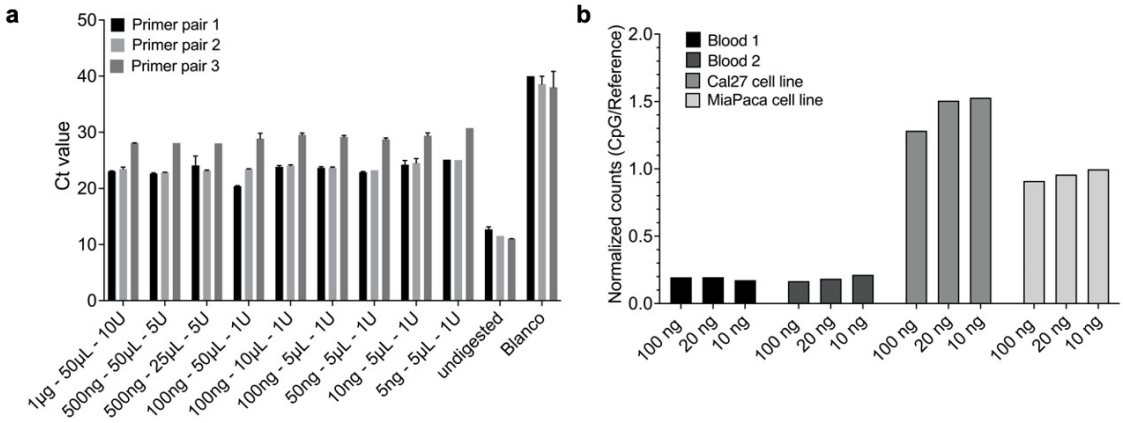
Suppl. Figure 3 | Percentage of non-digestion in digested samples based on spiked-in lambda phage DNA.



Suppl. Figure 4 | Density plot of the mean cross validated AUC (cvAUC) values for each single smMIP model. The distribution of cvAUC values is centered around 0.8 and is left-skewed. It is clear that the majority of smMIP models can make an accurate prediction.



Suppl. Figure 5 | Correlation of cvAUC values of all double-tiled smMIPs**.** Selection of all CpG sites targeted by two smMIPs. For those CpG sites, cvAUC of one smMIP is in the X axis and the other in the Y axis. The pearson correlation coefficient is r = 0.668.



Suppl. Figure 6 | Determination of the amount of input DNA. a) Results of the qPCR with lambda DNA digested in different reaction conditions. Lambda DNA was digested in different sample conditions. These conditions are provided in the x-axis: the input amount of lambda DNA, the reaction volume of the digestion reaction, and the amount of enzyme units of each of the enzymes. The undigested samples were treated the same as the digested samples, except for the addition of the restriction enzymes. 1ng of all samples was amplified with four primer pairs, each hybridizing around one of the MSRE restriction sites. One primer pair failed and was excluded from the graph. The bars represent the average Ct value of two duplicate samples. b) Normalized counts for different amounts of input DNA. 100ng, 20ng and 10ng of genomic DNA from two blood samples and two cell lines (Cal27 and MiaPaca) was used as input for IMPRESS. Plotting was performed using GraphPad Prism.

Supplementary Tables

Suppl. Table 1 | Overview of the used cell lines.

Name	Cancer type
HT29	Colorectal
HCT 116	Colorectal
SK-HEP-1	Liver
Miapaca	Pancreas
Cal27	Head and neck

Suppl. Table 2 | NGS quality parameters of NextSeq run.

Read	Cycles	(Projected) Yield (Gbp)	Aligned PhiX (%)	Error Rate (%)	Intensity cycle 1	%> Q30 (Quality)	Lane	Cluster PF (%)	Density
Read 1	75	35.20	3.40	0.19	7274.73	91.87	**1**	86.71 ± 0,67	212 ± 5
Read 2	8	3.33	0	0	6699.25	93.87	**2**	86.50 ± 0.63	211 ± 8
Read 3	8	3.33	0	0	4527.20	90.02	**3**	86.74 ± 0.61	211 ± 9
Read 4	75	35.20	3.33	0.28	6833.01	91.86	**4**	86.40 ± 0,77	210 ± 10
Non-index reads total	150	70.41	3.36	0.24	7053.87	91.86			
Totals	**166**	**77.07**	**3.36**	**0.24**	**6333.55**	**91.87**			

PART IV | Validation of novel biomarkers & technology

CHAPTER 7

Novel blood-based screening method for colorectal and breast cancer using an IMPRESSive assay

(Work in progress)

Isabelle Neefs[1,2], Joe Ibrahim[1,2], Thomas Vanpoucke[1,2], Wiebren Tjalma[2,3], Dieter Peeters[4], Christof Vulsteke[5], Lieselot Croes[6], Sabine Tejpar[7], Heiko De Schepper[8], Niels Komen[9], Souad Bouhadan[8], Pieter Dewinter[8], Timon Vandamme[6], Marc Peeters[2], Guy Van Camp[1,2], Ken Op de Beeck[1,2]

1 Centre of Medical Genetics, University of Antwerp and Antwerp University Hospital, Prins Boudewijnlaan 43, 2650 Edegem, Belgium

2 Centre for Oncological Research Antwerp (CORE), University of Antwerp and Antwerp University Hospital, Universiteitsplein 1, 2610 Wilrijk, Belgium

3 Multidisciplinary breast clinic, Antwerp University Hospital, Drie Eikenstraat 655, 2650 Edegem, Belgium

4 Department of Pathology, Antwerp University Hospital, Drie Eikenstraat 655, 2650 Edegem, Belgium

5 Department of Oncology, AZ Maria Middelares Hospital, Buitenring-Sint-Denijs 30, 9000 Gent, Belgium

6 Department of Oncology, Antwerp University Hospital, Drie Eikenstraat 655, 2650 Edegem, Belgium

7 Laboratory for Molecular Digestive Oncology, Department of Oncology, KU Leuven, Oude Markt 13, 3000 Leuven, Belgium

8 Department of Gastroenterology, Antwerp University Hospital, Drie Eikenstraat 655, 2650 Edegem, Belgium

9 Department of Abdominal Surgery, Antwerp University Hospital, Drie Eikenstraat 655, 2650 Edegem, Belgium

ABSTRACT

Introduction

Early detection of colorectal cancer (CRC) and breast cancer (BRCA) remains challenging, with current screening methods like the fecal immunohistochemical test (FIT) and mammography having limitations in sensitivity and participation rates. DNA methylation has emerged as a promising biomarker for early cancer detection due to its role in gene regulation and cancer-related aberrations. Methylated circulating tumor DNA (ctDNA) shows potential as a non-invasive biomarker, but analysis faces challenges such as DNA degradation and low concentration in blood. We previously developed IMPRESS, a method that allows simultaneous analysis of multiple methylated regions. Using this technology, we now aimed to create screening assays for CRC and BRCA.

Methods

We leveraged publicly available 450K array datasets to find differentially methylated sites for BRCA. We compared whole blood, normal, in situ and invasive carcinoma tissue to identify potential methylation biomarkers. For CRC, we partially re-used already performed EPIC array analysis to select sites that were compatible with IMPRESS and re-performed some analysis with whole blood. The latter was used in view of applicability in liquid biopsies. For both CRC and BRCA, we created a two-step assay including a 'cancer detection' and an 'invasiveness detection' panel, where the first panel aims to discriminate cancerous samples against normal samples, and the second panel is used to distinguish precancerous/in situ lesions from invasive tumors.

Results

The CRC assay consists of 180 smMIPs and was used for evaluation of 25 CRC, 38 adenoma and 30 normal colon samples. The cancer detection panel reached 100% sensitivity and 100% specificity. The invasiveness detection panel achieved 80% sensitivity at 92% specificity, with an accuracy of 0.88. For BRCA, a total of 152 smMIPs was used to analyze 9 BRCA, 9 in situ, 9 normal breast samples. The cancer detection panel has a sensitivity of 88.9% for invasive tumors and 100% for in situ carcinomas compared to normal samples. The accuracy of the invasiveness detection panel was 0.77, with a sensitivity of 66.7% at 88.9% specificity.

Conclusion

Our preliminary results indicate high sensitivity and specificity for discriminating cancerous samples against normal samples and good accuracy for distinction between precancerous and invasive tumors. Further research will be needed to validate these panels in external datasets. Lastly, these results provide a solid basis for evaluation of these panels in plasma-derived liquid biopsies.

INTRODUCTION

Cancer remains one of the most common causes of death. Colorectal cancer (CRC) and breast cancer (BRCA) are amongst the most common and deadly cancers world-wide. Early detection is crucial to improve survival rates, yet both cancers often progress undetected until advanced stages [1–3]. Therefore, screening programs have been implemented in the EU [4,5]. Especially for CRC, screening has proven to reduce its incidence and mortality. The fecal immunohistochemical test (FIT) is mostly used, as it is a simple, non-invasive test. It detects occult blood in feces and reaches sensitivities around 69.2% for the detection of invasive carcinomas, at 96.8% specificity. Moreover, the detection sensitivity for precursor lesions (adenomas) is suboptimal (71% at 97% specificity) and participation rates are less than desired [4,6,7]. For BRCA, a mammography is mostly used for screening purposes. This minimally invasive method reaches high specificities (98%), but sensitivity varies depending on breast density. Moreover, participation rates are, like in CRC screening, suboptimal [5,8]. This highlights the need for novel screening modalities.

In recent years, epigenetic alterations, and particularly DNA methylation, has gained attention for its pivotal role in cancer. In healthy cells, DNA methylation regulates gene activity and maintains genomic stability. However, in cancer, aberrant methylation patterns occur, silencing tumor suppressor genes and promoting oncogenesis. Given its early occurrence and universality across cancers, DNA methylation emerges as a promising biomarker for early cancer detection [9].

In view of novel screening methods, methylated circulating tumor DNA (ctDNA) shows promise as a non-invasive biomarker for cancer detection. Studies indicate that methylation patterns detected in primary tumors can also be found in ctDNA extracted from plasma, offering a potential avenue for early diagnosis [10–12]. However, challenges exist in methylated (ct)DNA analysis. Bisulfite sequencing, the gold standard method, leads to DNA degradation, which is particularly a problem in fragmented samples like ctDNA. Additionally, the low concentration of ctDNA in blood poses a limitation, as only a fraction of the genome may be represented. When analyzing only a few genes, which has been done in previous studies with multiplex MSP/qMSP, false-negative results are likely to occur [12]. This highlights the need for the simultaneous analysis of multiple methylated regions to improve sensitivity.

Our research group previously developed a highly multiplexable targeted methylation detection method, addressing the need for improved sensitivity and specificity in cancer detection. This method, called IMPRESS (Improved Methylation Profiling using Restriction Enzymes and smMIP Sequencing), can combine thousands of biomarkers into a single assay. In this study, we aimed to design two assays, for CRC and BRCA, using the IMPRESS technology, and to validate it on both fresh frozen and liquid biopsies. With this, we aimed to create

a novel, minimally invasive screening assay for CRC and BRCA, that can be easily implemented in clinical routine diagnostics.

MATERIAL AND METHODS

Sample collection and processing

Control samples

As a negative control, we used previously collected whole blood samples (n=7) from healthy volunteers. Genomic DNA (gDNA) was extracted using a standard salting-out procedure. DNA was stored at 4°C until further use. As a positive control, four cancer cell lines were used. For CRC, the HT29 and HCT116 cell lines were provided by the Centre for Oncological Research (CORE, Antwerp). For BRCA, the SKBR3 cell line was provided by CORE and the MCF-7 cell line was already stored in-house. All cell lines were cultured according to the American Type Culture Collection (ATCC) standard protocols. gDNA was extracted using the Blood & Cell culture DNA Maxi Kit (Qiagen, Germany). DNA was stored at -20°C until further use. As an internal control, unmethylated lambda phage DNA was used (Thermo Fisher Scientific, USA).

Fresh frozen tissue

CRC and BRCA tumor tissue and matched normal adjacent tissue samples were retrospectively collected from the Antwerp University Hospital (UZA, Belgium). A total of 57 colorectal (27 tumor and 30 normal adjacent) and 27 breast (9 tumor, 9 in situ and 9 normal adjacent) tissues were obtained and stored at -80°C. Colorectal adenoma fresh frozen tissues were collected through the Belgian Virtual Tumor bank network. A total of 38 samples were collected from 7 biobanks (see acknowledgements). Haematoxylin-eosin-stained sections per specimen were microscopically examined by a pathologist (D.P) to verify tissue type, presence of invasive tumor and overall tumor cell percentage (TcP). A minimum of 25% TcP was used for all analyses, with a median TcP of 40% and a range from 25% to 70%. Samples from all invasive cancer stages (I-IV) were used. DNA was extracted using the QIAamp DNA Micro Kit (Qiagen, Germany) according to the manufacturer's protocol. The DNA was stored at -20°C until further use. An overview of all samples is given in Table 1.

Table 1| Overview of collected samples

Tissue Type	Group	Number FF
Colorectal	Normal	30
	Adenoma	38
	Invasive carcinoma	25

	Normal	9
Breast	In situ	9
	Invasive carcinoma	9

FF= fresh frozen

Ethical Approval

Fresh frozen tissue samples were retrospectively collected from the Antwerp University Hospital Biobank. According to Article 20 of the Belgian Law on the procurement and use of human corporal material intended for human application or scientific research of 19 December 2008, patients provide consent for the use of their bodily material in research when consenting to an invasive procedure. As such, no additional consent was asked for the use of these retrospective samples. Prospectively collected blood samples were only collected after written informed consent was given by the subject. The study protocol and any modifications were approved by the UZA ethical committee (Ref. N°20/02/056 and Ref. N°41/14/426) before experimental analyses were performed.

Biomarker assay development

Target site discovery

For the discovery of potential DNA methylation biomarkers, publicly available methylation datasets were used. For CRC, a large part of the analyses was previously described [13]. For this study, the EPIC array datasets of carcinoma and adenoma were additionally compared to EPIC whole blood datasets as normal controls (Suppl. Table 1). Pairwise comparisons between β-values from different groups, being carcinoma, adenoma and whole blood were performed to find differentially methylated probes that overlap between normal adjacent samples and whole blood as compared to adenomas and carcinomas. (DMPs, see Table 2). For BRCA, 450K methylation data of invasive, in situ, normal and whole blood samples were used (Suppl. Table 1). Processing of the datasets was done as described before by Ibrahim *et al.* [14,15]. Pairwise comparisons between β-values of the different groups were performed to obtain DMPs (Table 2). Lastly, targets described in epithelial-mesenchymal transition (EMT) were searched in literature as EMT is known to play an important role in the in situ to invasive transition in BRCA [16–28].

Biomarker selection and assay development

After filtering DMPs on p-value (0.05) and restriction sites for at least one enzyme, several other criteria were set to find the most optimal biomarkers. For both CRC and BRCA, a two-step approach for building the biomarker assay was used. First, a panel for discriminating between normal and cancerous tissue was made. This is referred to as the 'cancer detection panel' Then, a panel for pre-cancer vs invasive carcinoma was constructed. This is

referred to as the 'invasiveness detection panel'. Figure 1 gives an overview of the experimental set-up. Table 2 gives an overview of the number of selected biomarkers per panel.

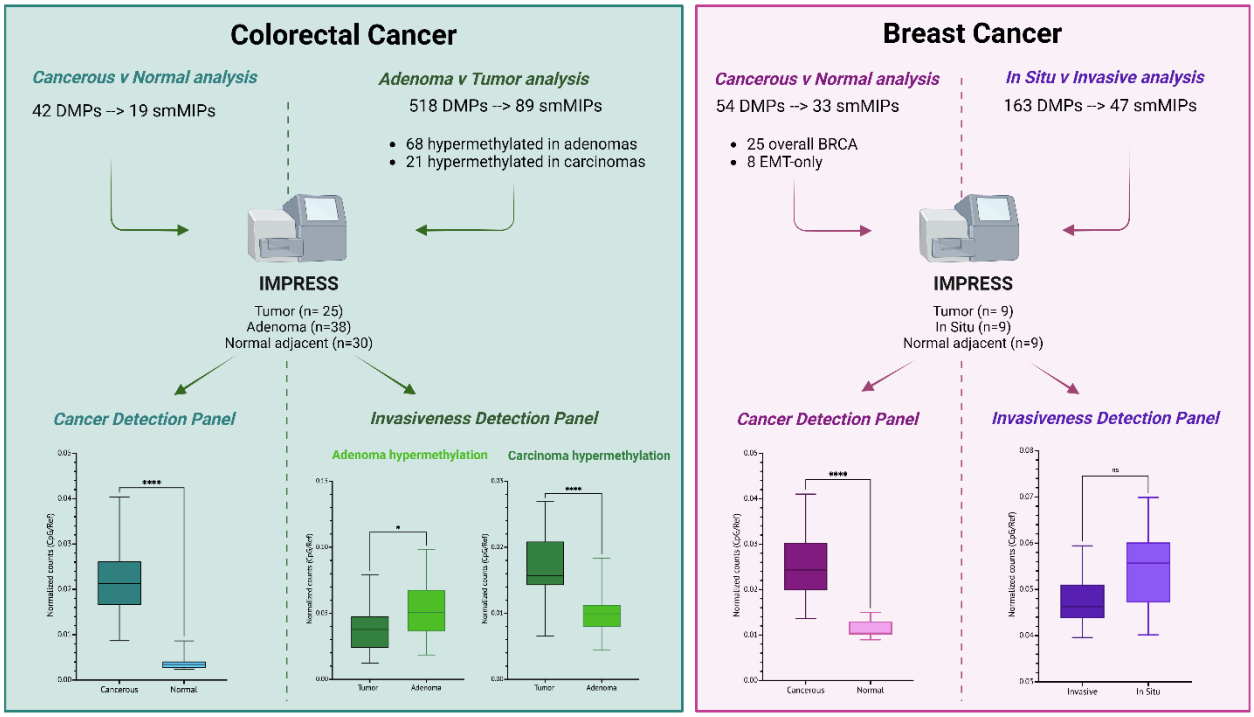


Figure 1 | Overview of the study. DMP= Differentially methylated probe

For CRC, the targets in the cancer detection panel were selected using following criteria: a) a minimal Δβ of 0.10 b) a maximal β-value of 0.35 in blood and normal adjacent samples and C) a minimal β-value of 0.45 in tumor. A total of 42 DMPs was selected. For the invasiveness detection panel, targets were selected based on following conditions: a) a minimal Δβ of 0.25 with maximally β= 0.27 in carcinomas and minimally 0.5 in adenomas or b) a minimal Δβ of 0.15 with β= 0.3 maximally in adenomas and β= 0.40 minimally in carcinomas. This led to 452 and 53 DMPs respectively. In our previous study [13], we also compared adenoma data to carcinoma data. The 13 DMPs that were found, were also used as targets for the subsequent smMIP design.

For BRCA, the targets in the cancer detection panel were selected as following: a) a minimal Δβ of 0.16 b) a maximal β-value of 0.54 in blood and normal samples and C) a minimal β-value of 0.2 in tumor. A total of 25 DMPs was selected. For the EMT targets, the conditions were the following: a) a minimal Δβ of 0.16 b) a maximal β-value of 0.20 in blood and normal samples and C) a minimal β-value of 0.22 in tumor. A total of 8 DMPs was selected.

For the invasiveness detection panel, targets were selected based on following criteria: a) a minimal Δβ of 0.16 with maximally β= 0.51 in invasive tumors and minimally β= 0.60 in in situ tumors or b) a minimal Δβ of 0.16 with β= 0.34 maximally in in situ tumors and minimally β= 0.37 in invasive tumors. This led to 143 and 20 DMPs respectively. For the EMT targets, we selected as following: a) a minimal Δβ of 0.15 with maximally β= 0.57 in

invasive tumors and minimally β= 0.57 in in situ tumors or b) a minimal Δβ of 0.11 with β= 0.48 maximally in in situ tumors and minimally β= 0.52 in invasive tumors. This led to 4 and 2 DMPs respectively.

smMIP design and selection

CpG smMIPs for CRC and BRCA were designed using the MIPGEN software [29]. smMIPs were designed for both DNA strands (i.e., double-tiled) for all selected targets. smMIPs in our design contain a) a common backbone of 30 nt b) a single molecule tag of 5 nt on both sides of this backbone c) two binding arms of circa 20 nt, specifically designed per target with an insert size of 50 nt. Single molecule tags were different per smMIP copy, allowing filtering for PCR duplicates. To maximize the number of target regions in the selection, only the smMIP with the highest theoretical performance score was retained. As such, a single-tiled smMIP selection was obtained. Also, smMIPs covering SNPs and repeats were removed. Table 2 illustrates the number of smMIPs per group. Reference smMIPs and lambda phage smMIPs were used as previously designed (**Chapter 6**). To balance the number of reference smMIPs towards the panels, we used only 50 smMIPs (>< 600 in **Chapter 6**).

Table 2| Overview of (selected) DMPs

Tissue	Step	Adenoma versus normal	Carcinoma versus normal	Adenoma versus carcinoma
Colorectal	DMPs	292 200	268 253	237 909
	Selected targets	42		518
	Final smMIP selection	19		89
		In situ versus normal	**Invasive versus normal**	**In situ versus Invasive**
Breast	DMPs	186 040	178 867	120 219
	Selected targets	54		163
	Final smMIP selection	33		47

DMP= differentially methylated probe

IMPRESS technology

We previously described our newly developed IMPRESS technology, its protocol and sample analysis in detail (**chapter 6**).

For the statistical analyses, differences in average methylation levels between invasive, non-invasive, and normal adjacent samples within one tissue type were tested using the Mann-Whitney U test (two-sided). For all analyses, p-values lower than 0.05 were considered significant. ROC curves were generated to determine the sensitivity and specificity of our panel. All statistical tests were performed in R (version 4.0.2) or GraphPad Prism (version 10.0.0) for macOS, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com.

RESULTS

Colorectal cancer

Initial observations and panel evaluation

For the first analysis, a subset of tumor (n=5), adenoma (n=5) and normal adjacent (n=5) samples was used. The library was prepared as described in **chapter 6**. In this analysis, 19 smMIPs were used for the cancer detection panel and 89 smMIPs were included in the invasiveness detection panel. Results for the panels are shown in Figure 2.

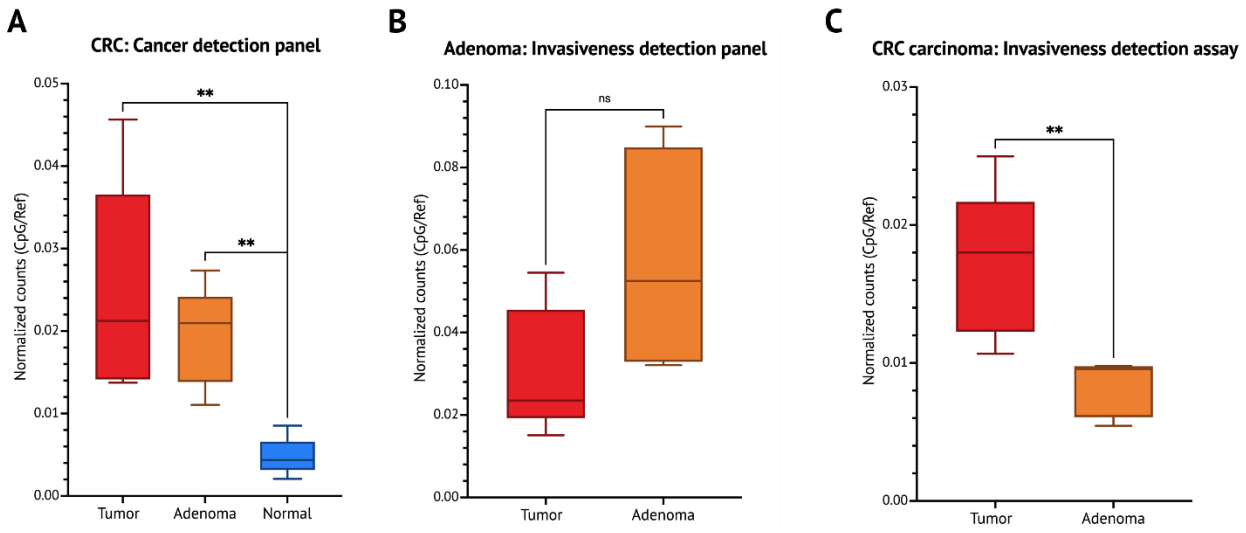


Figure 2 | CRC assay using n=5 per tissue type. A) Cancer detection assay B) Invasiveness detection assay with hypermethylated targets in adenoma C) Invasiveness detection assay with hypermethylated targets in carcinoma. Asterisks indicate p values where **= $p < 0.01$.

These results demonstrate that our selection for the cancer detection panel can effectively discriminate both cancerous and precancerous samples from normal samples (Fig. 2A). The results for the invasiveness detection panel show that targets for hypermethylation in adenomas (n=68) do not give a significant difference between precancerous and cancerous samples (Fig. 2B). However, looking at hypermethylated targets in tumors (n=21), thus different targets, in the same samples, a significant difference is observed (Fig. 2C).

Final CRC assay: diagnostic accuracy

Using the final CRC panels as described above, 25 tumor, 30 normal adjacent and 38 adenoma samples were analyzed with IMPRESS. Results of the cancer and invasiveness detection panels are summarized in Figure 3 and Table 3. To evaluate the diagnostic accuracy of both panels, ROC curves were built. The AUCs as well as

the sensitivity and specificity are reported in Table 3. ROC curves can be found in the supplementary material (Suppl. Fig. 1).

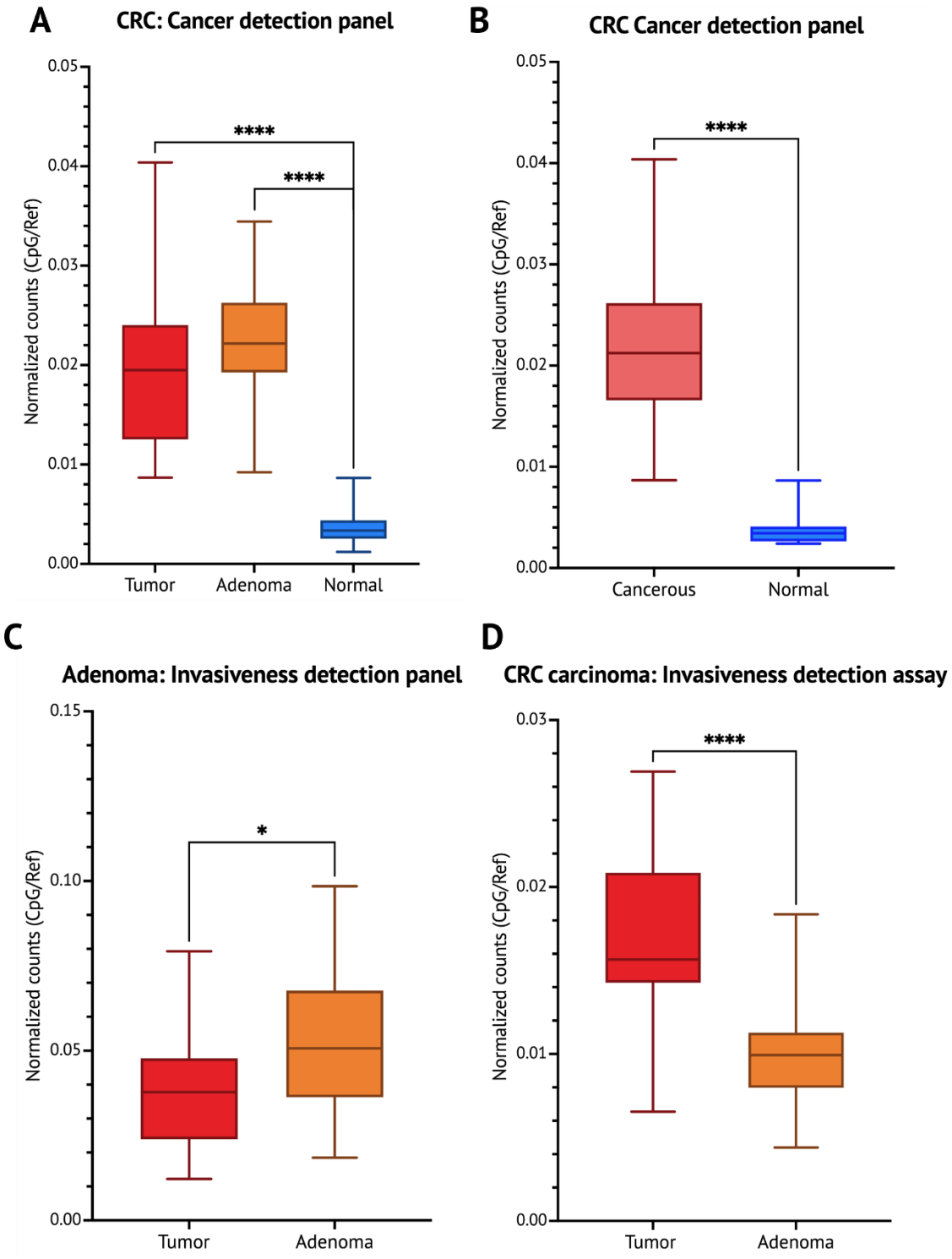


Figure 3 | CRC assay A) Cancer detection assay for each tissue type separate B) Cancer detection assay summary C) Invasiveness detection assay with hypermethylated targets in adenoma D) Invasiveness detection assay with hypermethylated targets in carcinoma. Asterisks indicate p values where *= $p < 0.05$ and ****= $p < 0.0001$

Table 3 | Summary of results and diagnostic accuracy for CRC panels

Panel	Difference in mean methylation level	Sensitivity	Specificity	AUC
Cancer detection panel: Tumor v Normal	****	100%	100%	1.00
Cancer detection panel: Adenoma v Normal	****	100%	100%	1.00
Cancer detection panel (Tumor + Adenoma v Normal)	****	**100%**	**100%**	**1**
Invasiveness detection panel: Tumor v Adenoma	****	80%	92%	0.88
Invasiveness detection panel: Adenoma v Tumor	*	47.3%	85%	0.69
Invasiveness detection panel (Tumor v Adenoma)	****	**80%**	**92%**	**0.88**

Hypermethylation is selected for the target group that is mentioned first. For example, in Tumor v Adenoma, targets are hypermethylated in tumor samples as compared to adenoma samples. P-values are indicated with asterisks, where *= $p < 0.05$ and ****= $p < 0.0001$.

Breast cancer

Initial observations and panel evaluation

A subset of invasive tumor (n=5), in situ tumor (n=5) and normal adjacent (n=5) samples was used in the first analysis. IMPRESS was used for library preparation and sequencing as described **chapter 6.** 33 smMIPs for the cancer detection panel and 47 smMIPs for the invasiveness detection panel were used. Results for the panels are shown in Figure 4.

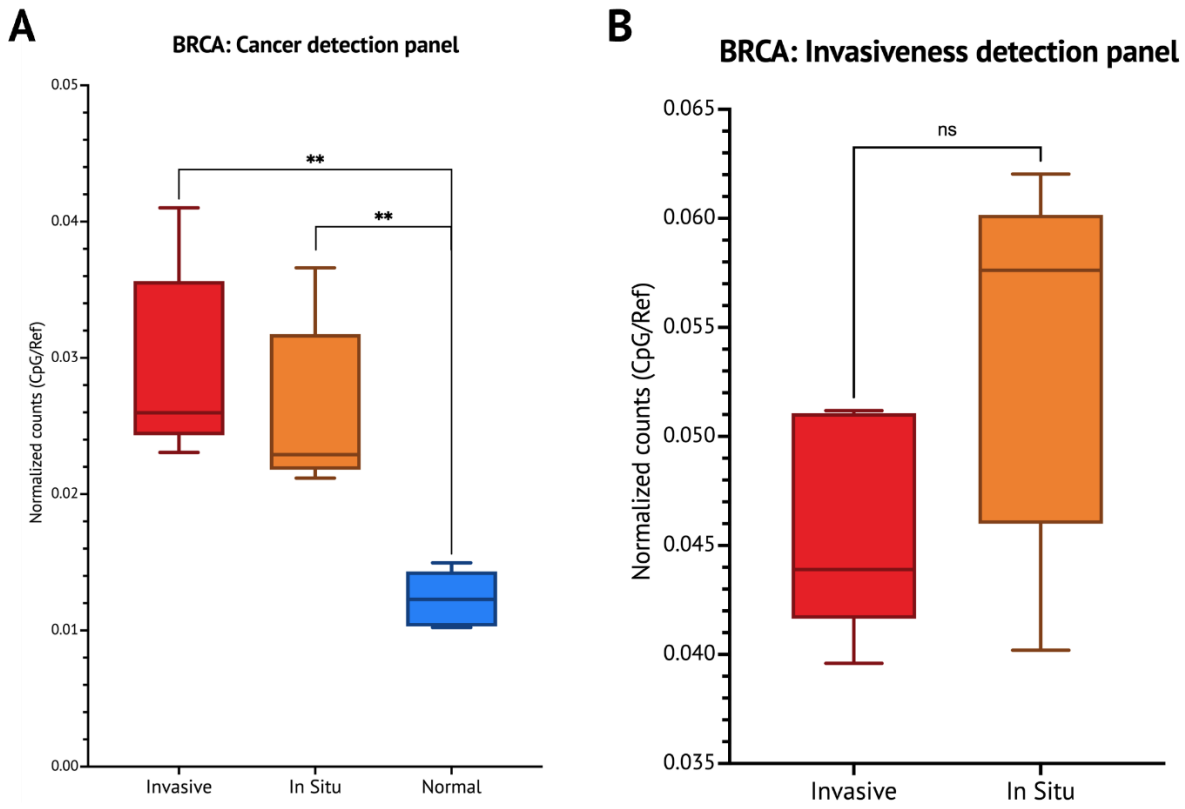


Figure 4 | BRCA assay using n=5 per tissue type. A) Cancer detection assay B) Invasiveness detection assay with hypermethylated targets in in situ. Asterisks indicate p values where **= $p < 0.01$.

These results show that the selected targets for the cancer detection panel can effectively discriminate (pre)cancerous samples from normal samples (Fig. 4A). The results for the invasiveness detection panel (Fig. 4B) show that no significant difference can be found between in situ and invasive samples (Figure 8, p-value = 0.09). smMIPs designed for hypermethylation in invasive samples compared to in situ samples, all failed.

Final BRCA assay: diagnostic accuracy

With the final BRCA panels, 9 samples of each group were analyzed. Results of the cancer detection and invasiveness detection panels are summarized in Figure 5 and Table 4. Assessment of the diagnostic accuracy

for both panels was done through ROC analysis. The AUCs, sensitivity and specificity are reported in Table 4. ROC curves can be found in the supplementary material (Suppl**.** Fig. **2**).

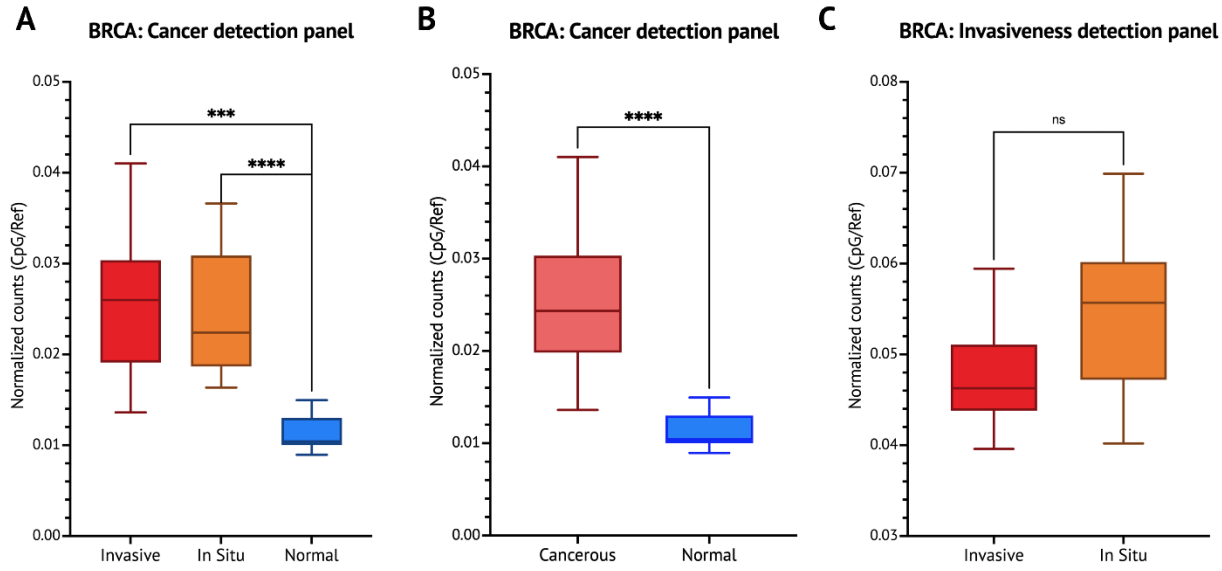


Figure 5 | BRCA assay A) Cancer detection assay for each tissue type separate B) Cancer detection assay summary C) Invasiveness detection assay with hypermethylated targets in in situ Asterisks indicate p values where ***= $p < 0.001$ and ****= $p < 0.0001$

Table 4 | Summary of results and diagnostic accuracy for BRCA panels

Panel	Difference in mean methylation level	Sensitivity	Specificity	AUC
Cancer detection panel: Tumor v Normal	***	88.9%	100%	0.98
Cancer detection panel: In Situ v Normal	****	100%	100%	1
Cancer detection panel (Tumor + In Situ v Normal)	****	**94.4%**	**100%**	**0.99**
Invasiveness detection panel : In situ v Invasive	p= 0.06	66.7%	88.9%	0.77
Invasiveness detection panel (In situ v Invasive)	**p= 0.06**	**66.7%**	**88.9%**	**0.77**

Hypermethylation is selected for the target group that is mentioned first. For example, in In Situ v Invasive, targets are hypermethylated in in situ carcinoma samples as compared to invasive carcinoma samples. P-values are indicated with asterisks, where ***= $p < 0.001$ and ****= $p < 0.0001$

DISCUSSION

In recent years, DNA methylation research has focused more on early detection and screening. These are very important parameters in cancer management, as early detection leads to better patient outcomes and higher survival chances. Therefore, population-based screening programs have been introduced. Despite this, disadvantages remain. We aimed to build novel DNA-methylation based assays using our IMPRESS technology, to create new sensitive and specific tests that can be easily implemented in the clinic. To this end, we have performed wetlab analyses to identify the most optimal biomarkers for our assay. In our *in silico* analysis, we included whole blood as a control in view of future liquid biopsy applications.

Although we are mostly interested in differentially methylated CpGs that make good potential biomarkers, it is interesting to retrospectively assess in which genes our selected sites are located (see Suppl Table 2 and 3). Regarding the CRC cancer detection panel, almost half of the genes we retrospectively found, are reported in literature (Suppl Table 2). Examples include *ADHFE1* and *OPLAH*, which we also reported in our previous study (**chapter 4**) [13]. Almost all targets have also been reported in the COSMIC database. For the invasiveness detection CRC panel, *RUNX3* has previously been reported by multiple research groups as an interesting target for adenoma detection through DNA methylation (see Suppl. Table 3). For example, Pasha *et al*. describe its potential as a serum biomarker for early CRC detection combined with *SFRP1* and CEA using MSP [30]. Remarkably, it has to our best knowledge not been further investigated in plasma samples, which are better suited for liquid biopsy purposes than serum samples. Looking at the hypermethylation targets in carcinomas, *SMAD4* and *UHRF1* overlap between our analyses and literature (Suppl. Table 4). The latter is involved in DNA methylation maintenance in CRC, making it potentially suited as a therapeutic target [31]. *SMAD4* is a known CRC gene, where previously mostly mutations have been described (see also Suppl Table 4 – Cosmic database) [32]. However, more recently, methylation of *SMAD4* has been described to play a role in metastasis in CRC [33]. Furthermore, it was described as a predictive biomarker [34]. These results once again underline the importance of DNA methylation as early detection targets in colorectal cancer.

Interestingly, from the 13 DE DMPs that we describe in our previous study (**chapter 4**) [13], only two could be included in the final assay. Starting from these 13 DE DMPS, six out of 13 targets were retained for the smMIP design due to the need for restriction sites for the MSREs. For these 6 targets, 5 smMIPs could be successfully designed. After analysis, ultimately two smMIPs performed well enough to be kept in the final (invasiveness detection) panel. The sensitivity and specificity for these two targets combined was 77% and 74% respectively, with an AUC of 0.81, which is lower compared to that of the 13 DE DMPs in **chapter 4** and compared to our final invasiveness detection panel (see Table 3).

As opposed to CRC, far less research has been performed on DNA methylation in BRCA (Suppl. Table 6-7). From the targets we selected for the cancer detection panel, only three have been reported in literature before. Contrary, most targets have been reported to the COSMIC database. *PITX2* has been reported on a few times a decade ago, where it has been described as a prognostic biomarker for BRCA. It was found both in primary tissue and blood serum samples [35,36]. The other two genes, *FAT4* and *SFRP2*, were selected in the analyses as EMT-specific targets as described in the method section. EMT is known to play a role in the carcinogenesis and metastatic potential of tumors. The process is modulated through various molecular pathways that can involve epigenetics [17,27]. In literature, methylation of *FAT4* and *SFRP2* were described in EMT-related research, although here only cell lines were used [20,24]. With our results, we preliminarily demonstrate that the methylation of these targets potentially plays an important role in BRCA-related EMT. Importantly, both genes have also previously been described to play a role in other cancer types such as ovarian cancer and CRC [37–40]. Interestingly, the only gene that has been previously described in literature that we also had in our target selection for the BRCA invasiveness detection panel, is the EMT-related *CDH1* gene. Loss of *CDH1* expression has been associated with progression from in situ to invasive BRCA [17]. Hypermethylation of this gene was reported as a prognostic biomarker [36,41]. Despite its presence in our panel, we must admit it is not our most ideal biomarker as the discrimination between in situ and invasive samples based solely on this target, reaches only a limited sensitivity of 33%. Further research on EMT–only targets and DNA methylation will be needed to obtain better performing biomarkers.

The diagnostic accuracy of the CRC cancer detection panel was evaluated using 25 tumor, 38 adenoma and 30 normal adjacent samples and a combination of 19 targets. With this set of samples, 100% sensitivity and 100% specificity could be found for both tumor and adenoma samples compared to normal adjacent samples. A similar sensitivity and better specificity were found compared to the results of our multi-cancer smMIP set of 358 smMIPs (**chapter 6**). For the tumor samples, a weak correlation (r= 0.5) could be found for the normalized count and tumor cell percentages. This is in line with results obtained using the multi-cancer smMIP set (**chapter 6**). In the future, cross-validation, analyses with larger sample sizes and validation in external datasets will be needed to estimate the performance of this biomarker panel more accurately. Nevertheless, these results are very promising.

For the invasiveness detection panel, the 25 tumor and 38 adenomas were analyzed using 68 targets with hypermethylation in adenomas and 21 targets with hypermethylation in tumors. The latter had the best performing diagnostic values with a sensitivity of 80% at 92% specificity, and an accuracy of 0.88. These results are a little worse than the accuracy we obtained in our previous study (**chapter 4**), however we used more samples in the current approach. Also, as mentioned before, we tried to include as many targets from the previous study in this experiment to validate them as biomarkers.

Despite the slightly worse performance, the CRC invasiveness detection panel gives quite promising results as compared to e.g. the *SEPT9* test, where 68-95% sensitivity at 80-90% specificity is reached for discriminating tumors from adenomas [42]. Further research will reveal if this performance holds when using cross-validation, external validation and liquid biopsies. Interestingly, targets with hypermethylation in tumors were better performing than targets with hypermethylation in adenoma, although these are more limited in numbers. In Fig. 3C, the difference between adenomas and carcinomas becomes significant as opposed to the first results (Fig. 2B). This could be due to the larger sample size, or due to the fact that a Mann-Whitney test is less strict than other statistical testing. More research is needed to discover whether this panel can be useful in the future. Lastly, in the current invasiveness detection panel, there is currently no distinction possible between adenoma-grade (low/high) nor adenoma type (e.g. tubulovillous, tubular, serrated…).

The diagnostic accuracy of the BRCA cancer detection panel was evaluated using 9 samples per subtype (invasive carcinoma, in situ carcinoma and normal adjacent samples) and a set of 33 smMIPs. Since literature suggests an important role for EMT in BRCA, we also verified the samples with the EMT smMIP set of 6 targets. Comparing in situ versus normal samples, there is no difference in performance of the overall BRCA cancer detection panel versus the EMT-only panel. However, comparing tumor versus normal samples, 88.9% and 77.9% sensitivity was found for the overall and EMT-only smMIPs respectively, at 100% specificity. Comparing these results to the multi-cancer smMIP set (**chapter 6**), we achieve comparable results with a slightly higher accuracy with this smMIP set. Interestingly, we achieve far better results with our overall panel compared to previous work by our group by Croes *et al*., who investigated 4 CpG sites in *DFNA5* [43]. There, a sensitivity of 61.8% at 100% specificity was found with an AUC of 0.83. This highlights the importance of combining multiple targets to increase sensitivity.

Contrary to the results of **chapter 6** and CRC, a good correlation is found between normalized counts and TcP with our BRCA specific panel (r = 0.82). This could be due to the limited range of TcP (40% -70%) that was used in these samples. Samples with lower TcP will be tested in the future, and results will unveil if the correlation will persist. Furthermore, cross-validation, analyses using larger and external datasets will be needed to estimate the performance of this biomarker panel more accurately. Nevertheless, these results are very promising.

In the invasiveness detection panel, the samples were compared using a subset of 47 smMIPs, all hypermethylated in in situ samples. In our analyses, no significant difference in methylation level was found. Nevertheless, a trend is observed (p= 0.06). This result could already be expected from the *in silico* analyses that were performed (see Suppl. Fig. 3), where the MDS plot indicates invasive and in situ samples have a strong overlap in beta values. Since in situ carcinomas often hold an invasive component and vice versa, it is quite difficult to reach high accuracy in this step.

Nonetheless, the AUC of 0.77 indicates that DNA methylation holds potential for discrimination of the two states. An important reason to make this distinction, is that not all in situ carcinomas will develop into an invasive carcinoma. More research will be needed to finetune our panel. Remarkably, all targets with hypermethylation in invasive carcinomas failed. In our CRC analysis, targets with hypermethylation in tumors give a better distinction, so this is potentially also the case for invasive v in situ BRCA samples. More research will be needed to verify whether more hypermethylation targets in invasive carcinomas can be found and whether they are indeed better performing than our current panel.

This study is a work in progress. A current restriction of this study is that we did not yet verify whether our targets are CRC or BRCA specific. For this, a more elaborate analysis against other cancer types will be needed. Also, it would be very important to check the targets for aberrant DNA methylation in e.g. inflammatory diseases such as IBD, Crohn and IBC. This can be done *in silico* using publicly available data or *in vitro* using clinical samples. For CRC, we already collected around 30 inflammatory colon liquid biopsy samples for future research.

In the future, the most important follow-up of this research is the validation of our panels in an external tissue dataset and ultimately in liquid biopsies. To this end, we already started collecting liquid biopsy samples (See suppl. Table 7). However, IMPRESS must first be optimized and validated for its use in liquid biopsies. To rapidly increase sample size, previously collected blood samples from the PANIB trial [44] and the lead-in FOLICOLOR trial [45] will be added. Moreover, samples collected in the COLIPAN (Collection of Liquid biopsies in Pan-cancer patients) project will also be included to increase sample size. Furthermore, an interesting future addition to our panels will be known mutations (e.g. *KRAS, BRAF, RASSF1* and *BRCA*) for CRC and BRCA.

In conclusion, we have created new biomarker panels for detection of CRC and BRCA, potentially in early stages. Our first results indicate high sensitivity and specificity for discriminating cancerous samples against normal samples and good accuracy for distinction between precancerous and invasive tumors. Further research will be needed to validate these panels. Lastly, we provide a solid basis for examination of these panels in plasma-derived liquid biopsies.

ACKNOWLEDGEMENTS

First of all, we would like to thank all patients included in this project for their voluntary and generous contribution to the research. We would like to thank the responsibles from all participating biobanks from the BVT involved in this project: CHU Brugmann, CHU Liège, CHU Mont-Godinne, CHU Saint-Luc, Institut Jules-Bordet, UZ Brussel and UZ Leuven.

REFERENCES

1. World Health Organisation and International Agency for Research on Cancer. GLOBOCAN - Cancer Today. 2020.

2. Colorectal Cancer: Statistics | Cancer.Net [Internet]. [cited 2024 Jan 25]. Available from: https://www.cancer.net/cancer-types/colorectal-cancer/statistics

3. Breast Cancer: Statistics | Cancer.Net [Internet]. [cited 2024 Jan 25]. Available from: https://www.cancer.net/cancer-types/breast-cancer/statistics

4. Labianca R, Nordlinger B, Beretta GD, Mosconi S, Mandalà M, Cervantes A, et al. Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2013;24:vi64–72.

5. Niell BL, Freer PE, Weinfurtner RJ, Arleo EK, Drukteinis JS. Screening for Breast Cancer. [cited 2024 Jan 25]; Available from: http://dx.doi.org/10.1016/j.rcl.2017.06.004

6. Issa IA, NouredDine M. Colorectal cancer screening: An updated review of the available options. World J Gastroenterol [Internet]. 2017 [cited 2024 Jan 25];23:5086–96. Available from: https://pubmed.ncbi.nlm.nih.gov/28811705/

7. Belgisch Kanker Register. Jaar rapport - Bevolkingsonderzoek naar kanker 2022. 2022.

8. Von Euler-Chelpin M, Lillholm M, Vejborg I, Nielsen M, Lynge E. Sensitivity of screening mammography by density and texture: A cohort study from a population-based screening program in Denmark. Breast Cancer Research [Internet]. 2019 [cited 2024 Jan 25];21:1–7. Available from: https://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-019-1203-3

9. Kulis M, Esteller M. DNA methylation and cancer. Adv Genet [Internet]. 2010/10/06. 2010;70:27–56. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20920744

10. Warton K, Mahon KL, Samimi G. Methylated circulating tumor DNA in blood: power in cancer prognosis and response. Endocr Relat Cancer. 2016;23:R157–71.

11. Moati E, Taly V, Didelot A, Perkins G, Blons H, Taieb J, et al. Role of circulating tumor DNA in the management of patients with colorectal cancer. Clin Res Hepatol Gastroenterol. 2018;42:396–402.

12. Elazezy M, Joosse SA. Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management. Comput Struct Biotechnol J [Internet]. 2018 [cited 2024 Jan 25];16:370. Available from: /pmc/articles/PMC6197739/

13. Janssens K, Neefs I, Ibrahim J, Schepers A, Pauwels P, Peeters M, et al. Epigenome-wide methylation analysis of colorectal carcinoma, adenoma and normal tissue reveals novel biomarkers addressing unmet clinical needs. Clin Epigenetics. 2023;15:111.

14. Ibrahim J, de Beeck KO, Fransen E, Peeters M, Van Camp G. The Gasdermin E Gene Has Potential as a Pan-Cancer Biomarker, While Discriminating between Different Tumor Types. Cancers (Basel) [Internet]. 2019 [cited 2024 Jan 25];11. Available from: /pmc/articles/PMC6896019/

15. Ibrahim J, Op de Beeck K, Fransen E, Peeters M, Van Camp G. Genome-wide DNA methylation profiling and identification of potential pan-cancer and tumor-specific biomarkers. Mol Oncol. 2022;16:2432–47.

16. Tomaskovic-Crook E, Thompson EW, Thiery JP. Epithelial to mesenchymal transition and breast cancer. Breast Cancer Research. 2009;11.

17. Nickel A, Stadler SC. Role of epigenetic mechanisms in epithelial-to-mesenchymal transition of breast cancer cells. Translational Research. Mosby Inc.; 2015. p. 126–42.

18. Ruike Y, Imanaka Y, Sato F, Shimizu K, Tsujimoto G. Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. BMC Genomics. 2010;11.

19. Das L. Epigenetic alterations impede epithelial-mesenchymal transition by modulating centrosome amplification and Myc/RAS axis in triple negative breast cancer cells. Sci Rep. 2023;13.

20. Le AVP, Szaumkessel M, Tan TZ, Thiery JP, Thompson EW, Dobrovic A. DNA methylation profiling of breast cancer cell lines along the epithelial mesenchymal spectrum—Implications for the choice of circulating tumour DNA methylation markers. Int J Mol Sci. 2018;19.

21. Khaled N, Bidet Y. New insights into the implication of epigenetic alterations in the EMT of triple negative breast cancer. Cancers (Basel). 2019;11.

22. Isert L, Mehta A, Loiudice G, Oliva A, Roidl A, Merkel OM. An In Vitro Approach to Model EMT in Breast Cancer. Int J Mol Sci. 2023;24.

23. Liu G, Liu YJ, Lian WJ, Zhao ZW, Yi T, Zhou HY. Reduced BMP6 expression by DNA methylation contributes to EMT and drug resistance in breast cancer cells. Oncol Rep. 2014;32:581–8.

24. Hernández-Oliveras A, Zarain-Herzberg Á. Expression and associated epigenetic mechanisms of the Ca2+-signaling genes in breast cancer subtypes and epithelial-to-mesenchymal transition. J Cell Commun Signal. 2022;16:461–74.

25. Carmona FJ, Davalos V, Vidal E, Gomez A, Heyn H, Hashimoto Y, et al. A comprehensive DNA methylation profile of epithelial-to-mesenchymal transition. Cancer Res. 2014;74:5608–19.

26. Lin YC, Lee YC, Li LH, Cheng CJ, Yang RB. Tumor suppressor SCUBE2 inhibits breast-cancer cell migration and invasion through the reversal of epithelial-mesenchymal transition. J Cell Sci. 2014;127:85–100.

27. Ribatti D, Tamma R, Annese T. Epithelial-Mesenchymal Transition in Cancer: A Historical Overview. Transl Oncol. Neoplasia Press, Inc.; 2020.

28. Lima JF, Nofech-Mozes S, Bayani J, Bartlett JMS. EMT in breast carcinoma—A review. J Clin Med. MDPI; 2016.

29. Boyle EA, O'Roak BJ, Martin BK, Kumar A, Shendure J. MIPgen: Optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics. 2014;30:2670–2.

30. Pasha HF, Radwan MI, Yehia AM, Toam MM. Circulating methylated RUNX3 and SFRP1 genes as a noninvasive panel for early detection of colorectal cancer. Eur J Gastroenterol Hepatol. 2019;31:1342–9.

31. Kong X, Chen J, Xie W, Brown SM, Cai Y, Wu K, et al. Defining UHRF1 Domains that Support Maintenance of Human Colon Cancer DNA Methylation and Oncogenic Properties. Cancer Cell. 2019;35:633-648.e7.

32. Pearlman R, Frankel WL, Swanson B, Zhao W, Yilmaz A, Miller K, et al. Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. JAMA Oncol. 2017;3:464–71.

33. Liu A, Yu C, Qiu C, Wu Q, Huang C, Li X, et al. PRMT5 methylating SMAD4 activates TGF-β signaling and promotes colorectal cancer metastasis. Oncogene. 2023;42:1572–84.

34. Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. Int J Mol Sci. 2017;18.

35. Göbel G, Auer D, Gaugg I, Schneitter A, Lesche R, Müller-Holzner E, et al. Prognostic significance of methylated RASSF1A and PITX2 genes in blood- and bone marrow plasma of breast cancer patients. Breast Cancer Res Treat. 2011;130:109–17.

36. de Ruijter TC, van der Heide F, Smits KM, Aarts MJ, van Engeland M, Heijnen VCG. Prognostic DNA methylation markers for hormone receptor breast cancer: a systematic review. Breast Cancer Res. 2020;22:13.

37. Duan H, Yan Z, Chen W, Wu Y, Han J, Guo H, et al. TET1 inhibits EMT of ovarian cancer cells through activating Wnt/β-catenin signaling inhibitors DKK1 and SFRP2. Gynecol Oncol. 2017;147:408–17.

38. Yang Q, Huang T, Ye G, Wang B, Zhang X. Methylation of SFRP2 gene as a promising noninvasive biomarker using feces in colorectal cancer diagnosis: a systematic meta-analysis. Sci Rep. 2016;6:33339.

39. Wei R, Xiao Y, Song Y, Yuan H, Luo J, Xu W. FAT4 regulates the EMT and autophagy in colorectal cancer cells in part via the PI3K-AKT signaling axis. Journal of Experimental & Clinical Cancer Research. 2019;38:112.

40. Malgundkar SH, Burney I, Al Moundhri M, Al Kalbani M, Lakhtakia R, Okamoto A, et al. FAT4 silencing promotes epithelial-to-mesenchymal transition and invasion via regulation of YAP and β-catenin activity in ovarian cancer. BMC Cancer. 2020;20:374.

41. Jung SP, Kim S, Nam SJ, Kim I, Bae JW. The Role of the *CDH1* Promoter Hypermethylation in the Axillary Lymph Node Metastasis and Prognosis. J Breast Cancer. 2013;16:16.

42. Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut. 2014;63:317–25.

43. Croes L, de Beeck KO, Pauwels P, Vanden Berghe W, Peeters M, Fransen E, et al. DFNA5 promoter methylation a marker for breast tumorigenesis. Oncotarget. 2017;8:31948–58.

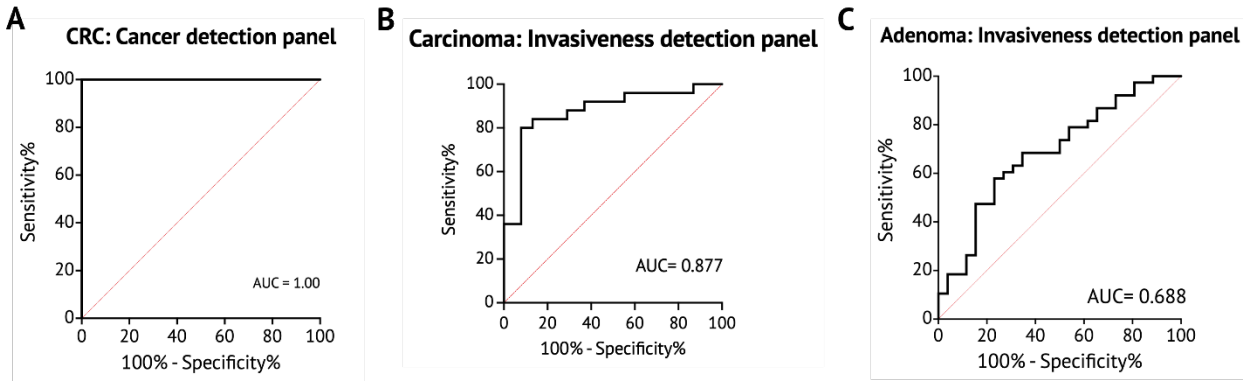
44. Janssens K, Vanhoutte G, Lybaert W, Demey W, Decaestecker J, Hendrickx K, et al. NPY Methylated ctDNA is a Promising Biomarker for Treatment Response Monitoring in Metastatic Colorectal Cancer. Clin Cancer Res [Internet]. 2023 [cited 2024 Mar 14];29:1741–50. Available from: /clincancerres/article/29/9/1741/725857/NPY-Methylated-ctDNA-is-a-Promising-Biomarker-for

45. University Hospital A. NCT04735900: lead-in FOLICOLOR Trial: Following Therapy Response Through Liquid Biopsy in Metastatic Colorectal Cancer Patients. https://clinicaltrials.gov/study/NCT04735900. 2022.

SUPPLEMENTARY MATERIAL

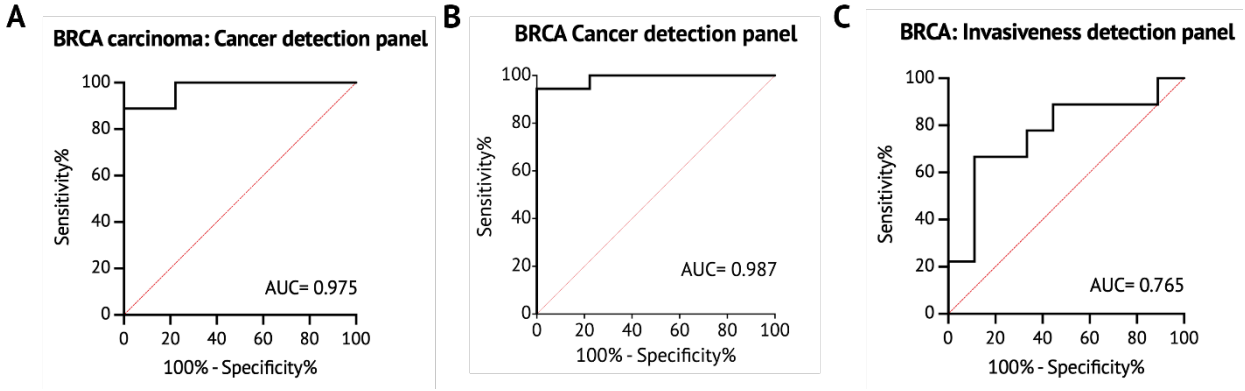
Supplemental Figures

Colorectal Cancer

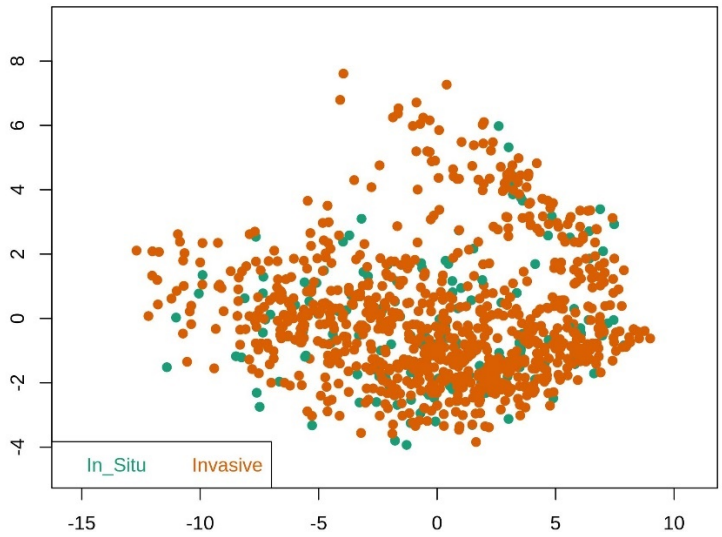


Supplemental Figure 1 | ROC plots of CRC assay. A) Cancer detection panel B) Invasiveness detection panel: hypermethylation in carcinoma C) Invasiveness detection panel: hypermethylation in adenoma. AUC= Area Under the Curve

Breast Cancer



Supplemental Figure 2 | ROC plots of step I and step II panel. A) First step panel (Invasive v Normal) B) First step panel (In Situ v Normal) C) Second step panel: In Situ v Invasive D) First step panel (Invasive v Normal), EMT only F) First step panel (In Situ v Normal), EMT only . AUC= Area Under the Curve



Supplemental Figure 3 | Multidimensional scaling plot of in situ v invasive breast cancer *in silico* analysis. The 1000 most variable positions are plotted. Publicly available 450K array data was used. Green dots = in situ samples (n= 155), orange dots = invasive samples (n= 1024).

Supplemental Tables

Supplemental Table 1 | Overview of datasets used for target discovery

Dataset ID - Colorectal	Adenoma	Carcinoma	Normal
E-MTAB-6450	16	0	0
E-MTAB-7854	80	0	0
GSE166212	10	32	0
GSE199057	0	77	0
GSE110554	0	0	12
GSE112618	0	0	6
GSE112905	0	0	10
GSE149412	0	0	292
GSE167998	0	0	12
GSE180119	0	0	16
GSE197331	0	0	17
GSE200376	0	0	19
GSE217633	0	0	44
GSE218186	0	0	55
GSE219293	0	0	18
GSE220622	0	0	136
Sum	**106**	**109**	**637**
Dataset ID - Breast	**In situ**	**Invasive**	**Normal**
GSE66313	40	0	0
GSE60185	115	93	0
GSE226569	0	694	0
GSE52865	0	183	0
GSE72308	0	54	0
GSE40279	0	0	0
Sum	**155**	**1 024**	**656**

Green: EPIC datasets, blue: 450K datasets

Supplemental Table 2 | Overview of first step panel's targets for CRC

Step I panel	Target	COSMIC?	Pubmed?
1:44873004-44873093	*ADHFE1*	yes	https://pubmed.ncbi.nlm.nih.gov/?term=ADHFE1+AND+CRC+AND+METHYLATION
1:101004882-101004971	*OPLAH*	yes	https://pubmed.ncbi.nlm.nih.gov/?term=OPLAH+AND+CRC+AND+methylation&sort=date
2:139537130-139537219	*NXPH2*	yes	/
2:61372147-61372236	/	/	/
2:182322224-182322313	*MDFI*	yes	https://pubmed.ncbi.nlm.nih.gov/?term=MDFI+AND+CRC+AND+METHYLATION
2:68546563-68546652	*COL4A1/COL4A2*	Yes – yes	https://pubmed.ncbi.nlm.nih.gov/24485021/
4:110223945-110224034	*CHST2*	yes	/
5:22853102-22853191	*TRBJ2-6/7*	No – no	/
6:73331049-73331138	*PREX2*	yes	/
6:41606401-41606490	*CNRIP1*	yes	https://pubmed.ncbi.nlm.nih.gov/?term=CNRIP1+AND+CRC+AND+METHYLATION
8:67344491-67344580	*PPP1R16B*	yes	https://pubmed.ncbi.nlm.nih.gov/?term=PPP1R16B+AND+CRC+AND+METHYLATION
8:53852627-53852716	*KCNQ5*	yes	https://pubmed.ncbi.nlm.nih.gov/?term=KCNQ5AND+CRC+AND+METHYLATION
8:72756098-72756187	*DMRTA2*	yes	/
8:145106519-145106608	*EDNRB*	yes	https://pubmed.ncbi.nlm.nih.gov/?term=EDNRB+CRC+AND+METHYLATION
8:68864463-68864552	*TRH*	no	/
10:100993555-100993644	*TM6SF1*	yes	/
10:13933953-13934042	*GPR88*	yes	/
11:128564082-128564171	*KCNA3*	yes	https://pubmed.ncbi.nlm.nih.gov/32407802/
13:50706751-50706840	*TTYH1*	yes	/
13:78493142-78493231	*c2orf74*	yes	/

Supplemental Table 3 | Overview of second step panel's targets (hypermethylated in adenomas)

Step II panel: hyper adenoma	Target	COSMIC?	Pubmed?
1:161695589-161695678	*FCRLB*	yes	/
1:180203108-180203197	*LHX4*	yes	/
1:217309141-217309230	*ESRRG*	yes	https://pubmed.ncbi.nlm.nih.gov/31566019/
1:235814317-235814406	/		
1:25255774-25255863	*RUNX3*	yes	https://pubmed.ncbi.nlm.nih.gov/?term=RUNX3+CRC+AND+METHYLATION
1:26372424-26372513	*SLC30A2*	yes	/
1:26551726-26551815	/		
1:32930480-32930569	/		
1:45083138-45083227	*RNF220*	yes	/
1:47691116-47691205	*TAL1*	yes	/
1:4771132-4771221	*AJAP1*	yes	/
1:85359068-85359157	/		
10:118429608-118429697	*C10orf82*	yes	/

10:119302826-119302915	*EMX2*	yes	/
10:119303779-119303868	*EMX2*	yes	/
10:119304561-119304650	*EMX2*	yes	/
10:43725098-43725187	*RASGEF1A*	yes	/
10:8094049-8094138	*GATA3-AS1*	no	/
11:31832846-31832935	*PAX6*	yes	/
11:35441759-35441848	/		
12:45270305-45270394	*NELL2*	yes	/
12:48398235-48398324	*COL2A1*	yes	https://pubmed.ncbi.nlm.nih.gov/33858326/
12:49484129-49484218	*DHH*	no	/
12:54378766-54378855	/		
12:62583916-62584005	*TAFA2*	yes	/
15:60289515-60289604	/		
15:65669504-65669593	*IGDCC3*	yes	https://pubmed.ncbi.nlm.nih.gov/27049830/
15:68117531-68117620	*SKOR1*	yes	/
15:93631885-93631974	*RGMA*	yes	https://pubmed.ncbi.nlm.nih.gov/19303019/
16:85063720-85063809	*KIAA0513*	yes	/
17:47653260-47653349	/		
17:48046800-48046889	*DLX4*	yes	/
17:59476264-59476353	/		
18:5629004-5629093	*EPB41L3*	yes	/
19:15619264-15619353	/		
19:36347964-36348053	*KIRREL2*	yes	/
2:170218620-170218709	*LRP2*	yes	/
2:172952973-172953062	*DLX1*	yes	/
2:177027863-177027952	/		
2:237477117-237477206	*ACKR3*	yes	https://pubmed.ncbi.nlm.nih.gov/30337690/
2:241759348-241759437	*KIF1A*	yes	https://pubmed.ncbi.nlm.nih.gov/34373442/
2:27070719-27070808	/		
2:289354-289443	/		
2:5506198-5506287	/		
2:69240058-69240147	/		
20:52790006-52790095	*CYP24A1*	yes	https://pubmed.ncbi.nlm.nih.gov/23463632/
22:46366799-46366888	*WNT7B*	yes	/
3:179169522-179169611	/		
3:96532840-96532929	/		
4:30722210-30722299	*PCDH7*	yes	/
4:37245464-37245553	/		
4:42399338-42399427	/		
4:57397099-57397188	*THEGL*	yes	/
4:996024-996113	*IDUA*	yes	/
5:153857758-153857847	*HAND1*	yes	https://pubmed.ncbi.nlm.nih.gov/38689296/

6:148663566-148663655	/		
6:70576991-70577080	*COL19A1*	yes	/
6:7728818-7728907	*BMP6*	yes	https://pubmed.ncbi.nlm.nih.gov/25227796/
6:90121298-90121387	*RRAGD*	yes	/
7:126891257-126891346	*GRM8*	yes	/
7:155249873-155249962	/		
7:155252143-155252232	*EN2*	yes	https://pubmed.ncbi.nlm.nih.gov/35169223/
7:28893582-28893671	/		
7:39454095-39454184	*POU6F2*	yes	/
7:8473218-8473307	/		
1:178030188-178030277	/		
6:31831420-31831509	*SLC44A4*	yes	/

Supplemental Table 4 | Overview of second step panel's targets (hypermethylated in tumors)

Step II panel: hyper carcinoma	**Target**	**COSMIC?**	**Pubmed?**
1:109133165-109133254	*FAM102B*	yes	/
1:33219200-33219289	*KIAA1522*	yes	/
1:52017727-52017816	/		
1:66799355-66799444	*PDE4B*	yes	/
10:118652771-118652860	*SHTN1*	yes	/
10:126412214-126412303	*FAM53B*	yes	/
11:128389383-128389472	*ETS1*	yes	https://pubmed.ncbi.nlm.nih.gov/22735571/
12:88969413-88969502	*KITLG*	yes	https://pubmed.ncbi.nlm.nih.gov/26674205/
13:100310911-100311000	*CLYBL*	yes	/
14:21731575-21731664	*KNRNPC*	no	/
14:52778453-52778542	/		
14:62164724-62164813	*KIF1A*	yes	https://pubmed.ncbi.nlm.nih.gov/34373442/
15:59543647-59543736	*MYO1E*	yes	/
18:48609211-48609300	*SMAD4*	yes	https://pubmed.ncbi.nlm.nih.gov/?term=SMAD4+CRC+AND+METHYLATION
19:4912177-4912266	*UHRF1*	yes	https://pubmed.ncbi.nlm.nih.gov/?term=UHRF1+CRC+AND+METHYLATION
22:40575548-40575637	*TNRC6B*	yes	/
3:125230719-125230808	*SNX4*	yes	/
3:186733780-186733869	*ST6GAL1*	yes	/
4:108797414-108797503	*SGMS2*	yes	/
4:140778351-140778440	*MAML3*	yes	/
4:38806030-38806119	*TLR1*	yes	/

Supplemental Table 5 | Overview of first step panel's targets for BRCA

Step I panel	**Target**	**COSMIC?**	**Pubmed?**
1:151811332-151811421	*C2D4D*	no	/

1:44031802-44031891	*PTPRF*	yes	/
1:119527570-119527659	*TBX15*	yes	/
2:66666991-66667080	*BARHL2*	yes	/
2:19561444-19561533	*MEIS1*	yes	/
3:181441502-181441591	/		
4:111550651-111550740	*SOX2OT*	no	/
4:154710722-154710811	*PITX2*	yes	https://pubmed.ncbi.nlm.nih.gov/32005275/
4:126238769-126238858	*SFRP2*	yes	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6164039/
4:126238894-126238983	*FAT4*	yes	https://pubmed.ncbi.nlm.nih.gov/34762262/
4:154709822-154709911	*FAT4*	yes	/
4:154710334-154710423	*SFRP2*	yes	/
4:126238402-126238491	*SFRP2*	yes	/
4:154710484-154710573	*FAT4*	yes	/
5:3607033-3607122	*SFRP2*	yes	/
5:141931150-141931239	*SFRP2*	yes	/
5:87440227-87440316	/		
6:148689043-148689132	/		
6:10391396-10391485	/		
6:105400946-105401035	*SASH1*	yes	/
6:85482543-85482632	/		
6:1389301-1389390	/		
7:117119540-117119629	/		
8:145698608-145698697	*FOXF2*	yes	/
12:54446011-54446100	*CFTR*	yes	/
12:32292929-32293018	*KIFC2*	yes	
12:52401468-52401557	*HOXC4*	yes	/
12:54321278-54321367	*IFFO1*	yes	/
14:38092046-38092135	*IFFO1*	yes	/
14:61118691-61118780	*BICD1*	yes	/
14:57276199-57276288	*GRASP*	yes	/
16:12183915-12184004	/		
17:59530036-59530125	/		
19:12978296-12978385	/		
1:151811332-151811421	*OTX2*	yes	/
1:44031802-44031891	*SNX29*	yes	/
1:119527570-119527659	/		
2:66666991-66667080	*MAST1*	yes	/

Supplemental Table 6 | Overview of second step panel's targets (hypermethylated in in situ samples)

In situ BRCA	Target	COSMIC?	Pubmed?
11:46342299-46342388	*CREB3L1*	yes	/
16:68823665-68823754	*CDH1*	yes	https://pubmed.ncbi.nlm.nih.gov/?term=CDH1+AND+BRCA+AND+METHYLATION
1:169866260-169866349	/		
1:209980146-209980235	/		
1:28416463-28416552	/		
1:34088550-34088639	*CSMD2*	yes	/
1:44466541-44466630	*SLC6A9*	yes	/
1:53809674-53809763	/		
10:114116638-114116727	/		
11:116955819-116955908	*SIK3*	yes	/
11:623734-623823	*CDHR5*	yes	/
11:62528216-62528305	/		
11:65090801-65090890	/		
11:73661160-73661249	/		
13:113365425-113365514	*ATP11A*	yes	/
13:34411571-34411660	*RFC3*	yes	/
15:96886768-96886857	/		
16:1339765-1339854	/		
16:1465340-1465429	/		
16:31270803-31270892	/		
16:83967295-83967384	/		
17:43976742-43976831	*MAPT*	yes	/
17:7258003-7258092	/		
18:29170539-29170628	/		
19:12253855-12253944	*ZNF20*	yes	/
19:17621009-17621098	/		
19:24184894-24184983	/		
19:41168291-41168380	/		
19:49455282-49455371	/		
2:163139656-163139745	*IFIH1*	yes	/
21:40751062-40751151	/		
22:30789988-30790077	/		
3:128779436-128779525	/		
3:152327458-152327547	/		
3:36806148-36806237	/		
3:53036887-53036976	*SFMBT1*	yes	/
3:73159890-73159979	/		
4:78431944-78432033	/		
5:600768-600857	/		

6:11112053-11112142	*SMIM13*	yes	/
6:31862434-31862523	*EHMT2*	yes	/
6:33033107-33033196	*HLA-DPA1*	yes	/
6:4286259-4286348	/		
7:66370263-66370352	/		
7:93535685-93535774	*GNGT1*	yes	/
8:125053098-125053187	*FER1L6*	yes	/
8:28610735-28610824	*EXTL3*	yes	/

Supplemental Table 7 | Overview of liquid biopsy collection

Tissue Type	Group	Number LB
Colorectal	Normal	134
	Adenoma	16* + 114
	Invasive carcinoma	7
Breast	Normal	100
	In situ	6
	Invasive carcinoma	73

* For these samples, a matched fresh frozen sample has also been collected. LB= Liquid biopsy

Supplementary methods

Liquid biopsies

For colorectum analysis, blood samples were collected from people between 50 and 74 years old undergoing a colonoscopy for both diagnostic (+FIT) and other purposes. People were considered healthy when no other cancer-related co-morbidities were described in the medical history, nor inflammatory colon diseases were found during the colonoscopy (or previously). A total of 134 healthy people and 114 people with colon adenomas were included. To evaluate concordance between tissue and liquid biopsies, especially in our invasiveness detection panel, matched tissue and blood samples were collected from 20 patients with adenomas. For CRC, a total of 7 early stage cancer patients were included. Blood samples were collected before resection of the tumor.

For breast analysis, blood samples were collected at the UZA and the AZ Maria Middelares. Blood was collected before therapy, including resection, chemo, radio and targeted therapy. A total of 6 patients with an in situ carcinoma and 73 patients with an invasive carcinoma were included. For CRC and BRCA, blood was sampled using Cell-free DNA collection tubes (Roche Diagnostics, Switzerland). Blood was handled within 72h using a

2-step centrifugation protocol (low-speed, high-speed) as described by Sorber *et al*. [1]. Plasma was stored at -80°C until cfDNA extraction.

For the controls of BRCA, a subset (n=100) of a larger collection of healthy volunteers is used. In collaboration with the SGS Clinical Pharmacology Unit (Edegem, Belgium), blood samples from healthy volunteers (>18y) have been collected. People were considered healthy when no cancer-related nor inflammatory comorbidities were described in medical history. Blood samples were collected in EDTA tubes and handled within 4-6 h according to the centrifugation protocol as described by Sorber *et al*. [1]. Plasma was stored at -80°C until cfDNA extraction. cfDNA was extracted using the QiaSymphony (Qiagen, Germany) according to the manufacturer's protocol. The DNA was stored at -20°C until further use.

References

1. Sorber L, Zwaenepoel K, Jacobs J, De Winne K, Goethals S, Reclusa P, et al. Circulating Cell-Free DNA and RNA Analysis as Liquid Biopsy: Optimal Centrifugation Protocol. Cancers (Basel) [Internet]. 2019 [cited 2024 Mar 14];11. Available from: https://pubmed.ncbi.nlm.nih.gov/30935089/

Part V | General Discussion & Future Perspectives

GENERAL DISCUSSION

Colorectal cancer (CRC) and breast cancer (BRCA) remain leading causes of cancer-related morbidity and mortality worldwide despite established screening programs. Early diagnosis remains challenging due to limitations in existing techniques such as fecal immunochemical test (FIT) and mammographies. In recent years, DNA methylation has emerged as a promising biomarker for early detection in many cancer types, including CRC and BRCA. Together with this, a move towards liquid biopsies for cancer detection is observed. However, DNA methylation remains technically challenging. Moreover, the ideal biomarkers are yet to be found. Aiming to address these problems, we focused on exploring the methylome for novel biomarkers and implementing them in novel assays, using both conventional and novel technologies.

Methylation: The ideal biomarker?

As already extensively highlighted throughout this PhD thesis, DNA methylation has several advantages which makes it a very promising biomarker. This is further underlined in the numerous studies that have been and are being performed. For CRC screening, several FDA-approved assays exist (aforementioned *SEPT9* and Cologuard), highlighting the potential of DNA methylation as clinical biomarkers. However, they still need to find their way into the clinic, especially in the EU. One of the major hurdles here, is the lack of sensitivity for detection of advanced adenomas compared to colorectal carcinomas. The m*SEPT9* assay for example, only reaches a detection sensitivity ranging from 7.9% - 38.7% for adenomas [1–3]. Other blood-based assays for CRC detection are being developed, where a trend is seen towards combining multiple methylated biomarkers. For example, the ColoDefense assay combines *SEPT9* with Syndecan 2 (*SDC2)*, leading to an increased sensitivity of 47.8% for adenomas [1]. Although sensitivity remains quite low, it is clear that combining biomarkers is important for enhanced detection of adenomas [3].

Moving towards the ideal biomarkers for CRC screening and early detection, we have investigated DNA methylation targets in **chapter 4**. The results of this study are in line with the general idea that DNA methylation biomarkers are powerful targets for cancer detection, especially when several biomarkers are combined. Our most promising results were the 13 DMCs that can accurately discriminate adenomas from carcinomas, as this remains an important clinical issue. Future validation of these biomarkers, in both tissue and liquid biopsies will confirm if the promising *in silico* diagnostic accuracy can be maintained.

For BRCA, the best example regarding DNA methylation in diagnostics, is the TCGA analysis for triple negative breast cancer (TNBC). The profiling of the methylome of these carcinomas in the TCGA project has revealed specific DMRs associated to TNBC, which may serve as potential biomarkers for subtyping [4]. Again, combining different regions or sites was key to reach higher sensitivities.

Most published studies discussed above, investigated (multiple) methylation biomarkers using MSP or qMSP [5]. However, as discussed in **chapter 1**, these approaches only have limited multiplex capacities. For example, to validate our own results from **chapter 5,** we would need to use another technology as we have more markers than can be multiplexed in MSP/qMSP. Recent advances in methylation detection technologies have brought forward mostly targeted NGS approaches, but at high cost. Therefore, we have proposed a new method in **chapter 6**, which will be discussed below.

Towards improved assays for DNA methylation detection

DNA methylation is clearly an ideal biomarker candidate for early detection. As comprehensively highlighted in **chapter 1**, there exist many technologies for methylation detection. However, some of them are more suitable for development of a cancer screening assay compared to others. Ideally, screening must be performed in a non-invasive or minimally invasive manner, by using liquid biopsies. A liquid-biopsy based screening assay must be low-cost, highly sensitive (can be achieved e.g. by high multiplex capacity), high throughput, and have a low analytical detection sensitivity. The latter is especially important when using for example cfDNA, as the ctDNA fraction will be very low in early stages. As also described in **chapter 1**, most of the CE-IVD approved tests, are PCR-based. However, PCR-methods have some important disadvantages (see Table 1). Other techniques with better detection sensitivities, such as ddPCR and NGS are nowadays becoming more popular. Nevertheless, these also come with important limitations (see Table 1). For example, recent technologies such as TAPS and EM-seq have advantages in multiplex capacities and detection sensitivity but remain quite expensive. In that view, we have developed IMPRESS, as described in detail in **chapter 6**.

Table 1 | Overview of performance of different technologies in various aspects

Aspect	PCR methods	ddPCR	Other NGS-based techniques	IMPRESS
Cost	€	€	€€€	€€
Detection sensitivity	++	+++	+	+
Throughput	+/-	+/-	+	+
Turn-around time (TAT)	++	++	+/-	+
Multiplexing capacity for high diagnostic accuracy	-	-	+	+

With IMPRESS, we strived for developing a technology with all ideal assets. We have created a low-cost method (~ €20/sample), with a reasonable detection sensitivity that is currently still being optimized. Moreover, it is high throughput and with the smMIPs, many targets can be combined (>2000), leading to increased sensitivity. It is especially interesting to compare this with the results from **chapter 5,** where we used ddPCR. The cost for ddPCR is comparable (~ €15/sample) to IMPRESS, but only 96 samples can be analyzed together. One of the biggest technical differences, is that the ddPCR analyses were done using bisulfite conversion. This remains up

until now the gold standard approach in ddPCR. However, in view of liquid biopsies, it might be important to switch to another approach, e.g. an enzymatic approach, to avoid DNA loss. NGS has a bigger multiplex capacity than ddPCR, which can drastically increase sensitivity. Contrary, ddPCR has the major advantage that its detection sensitivity is extremely low. Theoretically, it can reach 0.001% detection sensitivity, but we observed sensitivities down to 0.8% in our study. This is still quite low compared to NGS technologies (1%-2%) [6,7]. With IMPRESS, the limit for methylation detection (LOD) is 4% but is currently being optimized to reach a lower LOD, which is necessary for low ctDNA values in early detection.

In view of performance, IMPRESS and ddPCR are very alike when it comes to biomarker analysis on fresh frozen tissue. In **chapter 5**, we developed 2 ddPCR assays encompassing a total of three targets, which were also analyzed using IMPRESS in **chapter 6**. Therefore, we can directly compare these two technologies. To be able to compare, a single model evaluating sensitivity and specificity was built for all different targets. For IMPRESS, ROC-curves were made based on the normalized counts. For ddPCR, the ROCs were built using the methylation levels. The sensitivity and specificity of the single ddPCR models were compared with those from the single smMIP models for the three targets. Table 2 shows that only minimal differences between both technologies are found. These results indicate that the IMPRESS technique performs at least equally well as the gold standard ddPCR.

Table 2 | Overview of target performance in ddPCR versus IMPRESS

Target	Sensitivity		Specificity		AUC	
	ddPCR	*IMPRESS*	*ddPCR*	*IMPRESS*	*ddPCR*	*IMPRESS*
chr2:73147755-73147844	83.5%	81.1%	98.2%	97.4%	0.916	0.898
chr5:76923876-76923965	81.6%	79.0%	88.2%	94.7%	0.861	0.913
chr7:8482030-8482119	72.3%	87.2%	91.8%	83.3%	0.842	0.892

Single versus multi cancer detection

From our results, we believe IMPRESS could become an important player in the DNA methylation detection field. Taken together with methylation as a rising biomarker in the clinic, our goal is to implement IMPRESS assays as early detection tools for cancer. In **chapter 6**, we have addressed this by building a biomarker panel covering eight of the most frequent cancer types, a so-called multi-cancer detection (MCD) assay. However, in **chapter 7**, we focused on single-cancer assays for CRC and BRCA. For this study, we used a different methylation signature than the one for our multi-cancer study. Several reasons explain this difference. First, our multi-cancer methylation signature was based on TCGA data, which only has 450K methylation array data. However, with the introduction of the novel EPIC methylation array, large EPIC datasets were built for some cancer types. As described in **chapter 4**, such EPIC datasets were publicly available for CRC. Moreover, we also built our own

in-house dataset using EPIC arrays. As such, a novel methylation signature for CRC was made. For BRCA, there were ample EPIC datasets, so we initially chose to expand the TCGA analysis with additional public datasets as we also wanted to investigate in situ carcinomas that are not present in the TCGA. After this search, we ended up with a large dataset, in which combining TCGA with this dataset would lead to a big imbalance for invasive carcinomas. Therefore, we focused only on the novel datasets that also included in situ carcinomas.

In Flanders, screening programs are only implemented for three cancer types (BRCA, CRC and cervical cancer). In this view, it is much more feasible to create a single cancer assay that can be directly compared to the current gold standard screening method. We did not focus on cervical cancer in this PhD thesis, as in the initial TCGA analyses, there were too little controls to allow for robust analyses [8]. Lastly, screening has only been proven effective in CRC, as both mortality and incidence have been reduced [9,10]. In other cancer types, for example breast cancer, benefits of screening remain debated [10–12]. Therefore, the introduction of the MCD in clinic is much more complex than a single-cancer assay.

Of course, there are also some disadvantages when using single-cancer assays. For example, many patients present with vague symptoms, such as weight loss, pain, tiredness… [13]. Screening for cancer with multiple single cancer tests requires performing these tests sequentially. However, as the false-positive rates (FPR) of the separate tests are cumulative, this approach is prone to high FPRs, which in turn can lead to an increased number of unnecessary, invasive follow-up procedures and stress for the patient. In that light, MCDs have gained much more attention in recent years.

The holy grail: Multi-cancer early detection

Given the challenges that exist for current screening methods, a blood-based test for simultaneous detection of multiple cancer types in an early stage (multi-cancer early detection, MCED) has become an important goal in oncological research [14]. There are some important advantages why MCED could prove very useful in clinic [15,16]. As mentioned before, using single cancer tests sequentially leads to high FPR. In contrast, the reported false-positive rate for most MCEDs is <1% [16]. Furthermore, in the discussions about cancer screening, there has been suggested that broad coverage of cancer by screening, in particular blood-based screening, could potentially deliver the greatest benefit in patients. Especially for cancers that are considered curable and can progress quickly, such as BRCA, CRC but also lung, esophagus and head & neck cancer, early detection is considered beneficial [10]. Also, current screening programs only cover 30% of all cancer-related deaths among the eligible screening population. MCEDs can become a screening option for 70% of lethal cancers that have currently no effective tests. And for the cancers that do have screening options, MCEDs could be used to find cancers in people that are ineligible for - or missed by standard screening [17]. This has been described in the study of Hathaway *et al.* They found that an additional 11% of women could have been screened should an

MCED be present during a routine examination[18]. Lastly, an MCED would be a more cost-effective approach than single-cancer assays, especially for low prevalent cancers [14,16,19].

Next to all health-related advantages of MCEDs, a crucial aspect to consider is patient preference regarding such tests. In that view, Myers *et al.* conducted a study in older adult primary care patients, to assess their receptivity to an MCED test [20]. A total of 159 patients were included. They found that the interest in MCED testing was high (79%) and was not influenced by amongst others: age, gender, ethnicity and education level. Another study by Gelhorn *et al.* showed comparable results, where 72% of respondents (n=1700) preferred MCED screening to no cancer screening [21]. However, both studies were done in the USA, so European studies are needed to verify whether patient preference might differ in this different demographic background.

A widely debated disadvantage of cancer screening, which will be particularly important in MCED, is overdiagnosis and subsequent overtreatment. This occurs upon detection of a cancer that would not have progressed to impact life expectancy nor quality of life. Intuitively, this is more challenging in MCED than in single cancer assays, as multiple cancers ranging in aggressiveness are detected with an MCED. In the Hippocratic Oath, health providers swear to 'do no harm;' However, overdiagnosis can lead to either physical harm and/or psychological harm. It is of extreme importance that MCEDs will be evaluated in further studies to understand the implications of MCED screening compared to standard (single cancer) screening methods and to determine the ideal MCED testing interval to balance cancer detection and overdiagnosis [17,22,23]. An important trial in this aspect will be the Vanguard study, discussed below.

Most MCEDs that have been developed, focus on the analysis of circulating molecules in blood and cfDNA. It was already outlined in the discussion of **chapter 6** that there exist multiple MCEDs (CancerSEEK, PanSeer, Galleri), each focusing on distinct types of modifications, different circulating molecules and using different technologies [22]. Nevertheless, many of the described assays and studies focus on DNA methylation, which is in line with our research regarding an MCED IMPRESS assay. To outline the importance of DNA methylation MCEDs and potential clinical impact, a few clinical studies will be discussed below.

One of the biggest companies focusing on MCED through DNA methylation is GRAIL. They have developed the aforementioned Galleri test, which can detect 50 cancer types through liquid biopsy. GRAIL has funded many large-scale clinical studies, amongst others the Circulating Cell-free Genome Atlas (CCGA, NCT02889978) [14]. The CCGA study included over 15 000 participants with and without cancer for longitudinal follow-up. The study was divided in three subparts: 1) to evaluate three different sequencing assays [16] 2) Chosen assay development, training and validation for simultaneous detection of 50 cancer types and tissue of origin (TOO) [24] and 3) Independent validation of the assay, specific for screening purposes [25]. In the first sub-study, methylation analysis (whole genome bisulfite sequencing) outperformed genomic technologies (WGS and targeted mutation analysis), emphasizing once again that DNA methylation analysis is the most promising

approach for early cancer detection [14,16]. In the second study, a targeted methylation assay was developed, trained and validated in over 6 500 participants. The goal here was to simultaneously detect over 50 cancer types and discriminate in TOO. Prediction accuracy for the latter was 93%, while a sensitivity of 43.9% for early detection was found at 99.3% specificity [14,24]. In the independent validation sub study, TOO prediction accuracy decreased to 88.7%, while sensitivity and specificity increased to 51.5% and 99.5% respectively [14,25,26]. Besides the CCGA study, GRAIL is conducting other trials to improve the test's potential as a screening tool, a.o. the PATHFINDER (NCT04241796) and NHS-Galleri trial (ISRCTN91431511). The PATHFINDER trial already enrolled over 6 500 asymptomatic participants in the USA for assessment of the Galleri test in clinical practice. Unique for this trial, is that test results are returned to health care providers and patients for medical intervention [14,17,26,27]. The NHS-Galleri trial is conducted in the UK. For this trial, asymptomatic individuals are being recruited to evaluate whether the MCED can reduce the incidence rate of late-stage cancer in the intervention compared to the control arm [27]. An important trial that is being launched, is the Vanguard trial from the National Cancer Institute (NCI). In this trial, MCED blood tests will be selected and assessed to determine their efficacy and also safety in 1) detecting cancer and 2) preventing cancer-related deaths [20,28]. To check for safety is an important goal because, in addition to how a test performs, harms associated with cancer screening (described above) must be evaluated. We can learn from these studies, as they might highlight both improvements and challenges to overcome in designing future large-scale studies that we aim to perform with IMPRESS.

Although GRAIL has done some remarkable work, there is still important added value of other MCED, such as the IMPRESS assay that we are currently developing for liquid biopsies. As already touched upon in table 1, IMPRESS has the advantage of being a low-cost technology with low-cost assays. Furthermore, IMPRESS has the advantage of being bisulfite-free, contrary to the Galleri test. In that view, we expect to achieve even better diagnostic accuracy. Of course, several important steps will first need to be taken before we can fully compare our IMPRESS MCED to GRAIL's. In this view, an essential and ongoing aspect of our research is the TOO prediction with IMPRESS. Our current assay is not able to distinguish cancer signals, which is of utmost importance for clinical use. Lastly, IMPRESS must be further optimized for use in liquid biopsies (currently ongoing) and later also validated in independent cohorts and large clinical trials.

Applications of IMPRESS beyond cancer detection

Although we developed IMPRESS with a biomarker panel intended for use in early cancer diagnosis, there are far more options for this technology. One very important future application in cancer, is minimal residual disease (MRD) detection in liquid biopsies. For this, the LOD of IMPRESS must be optimized to reach very low sensitivities. As IMPRESS is a low-cost technology with a relative short TAT, patients could be easily followed-up throughout their remission. Another potential application is therapy follow-up. Researchers within our group

are focusing on DNA methylation biomarkers for both intrinsic and acquired resistance to therapy. These patterns could then be analyzed using specific IMPRESS panels.

Besides applications in cancer, IMPRESS could be useful for implementation in any DNA methylation-related research field and disease. An important example would be detection of viral infections, such as SARS-CoV-2. Research throughout the pandemic has made clear that DNA methylation patterns are important predictors for COVID-19 severity in affected individuals and can discriminate patients from uninfected persons [29]. Furthermore, it became clear that cost-effective high-throughput methods were lacking. In this view, IMPRESS panels designed for detection of either viral methylation or host methylation due to viral infection could become an important added value in the infection disease field. Furthermore, IMPRESS could become an important asset in the neuroscience field. Especially for use of early detection of DNA methylation biomarkers related to aging, and thus disease in elderly such as dementia or other cognitive impairment conditions. Recent research has highlighted the role of DNA methylation in these diseases and the potential of early detection methylation biomarkers [30,31]. Lastly, epigenetic research beyond biomedical applications, for example biological or biosocial research topics, can also benefit from our technology [32]. In all, IMPRESS has many potential applications, both in a research setting as well as in future clinical implementation.

Liquid biopsies: current contributions and challenges

One of the main reasons we developed IMPRESS, is to have a low-cost, highly sensitive and bisulfite-free technology for cancer detection in cfDNA. Throughout this PhD, IMPRESS was developed and optimized for use in fresh frozen tissue (**Chapter 6**). We have included a few preliminary results of liquid biopsies in this chapter, where we already highlight the feasibility of using IMPRESS for cancer detection in liquid biopsies. Besides these results, we have continued to optimize IMPRESS for use in cfDNA. Currently, one of our main challenges is the low amount of cfDNA – and thus ctDNA that is present in samples, which is a known problem in the field. We have successfully sequenced down to 5ng of cfDNA, but some samples are not reaching this input amount. A reason for this might be that we use leftovers from the NIPT samples. To take up the challenge of sequencing lower input amounts, we have already tried several approaches. We started with sequencing lower input amounts of 1-3ng cfDNA, but we noticed that too many so-called 'empty smMIPs' are created. We already established that changing the smMIP-to-DNA ratio does not affect this. It will be further investigated by other researchers of our group. Furthermore, we looked into ways of increasing the input amount of the already extracted samples. A first strategy we used, was vacuum centrifugation of the cfDNA samples that were already extracted. Unfortunately, this increased the salt concentration too much for the restriction enzymes to digest the samples properly. We also used bead purification, to lower the volume of sample and as such increase the concentration. Regrettably, too much DNA is lost with this procedure. Lastly, we looked into changing the elution volume upon cfDNA extraction. This is of course only an option for plasma samples where cfDNA has

not been extracted yet. This is currently ongoing, and up until now the most promising method to increase the cfDNA input amount. In literature, plasmapheresis has been suggested as a potential strategy to obtain more cfDNA [33]. However, this could only be used in e.g. minimal-residual disease detection and is less favorable for screening purposes.

Throughout this PhD, several blood sample collections were started. Colorectal and breast samples were specifically collected for this PhD (see **chapter 7**). In view of future validation of IMPRESS, the COLIPAN (Collection Of Liquid biopsies in PANcancer patients) study was initiated. In this study, adults (>18 y/o) with a new cancer diagnosis are asked to participate by donating three blood samples before treatment. Included cancer types are colorectal, breast, lung, esophageal, head & neck, liver, pancreatic and prostate cancer. In this study, over 60 patients have been included. Furthermore, we also looked into collecting blood samples from people with other conditions where plasma cfDNA levels are elevated, e.g. auto-immune diseases (n=20) and inflammatory diseases (n=53). This is important to make sure we have a cancer-specific assay. Lastly, blood samples from healthy volunteers are being collected in collaboration with SGS, a company focusing on clinical trials. Here, over 300 people have been included, with a total of more than 150 persons aged older than 50. As such, we have age-matched controls for our cancer patient samples.

One of the main difficulties in liquid biopsy research, is the implementation of pre-analytical standards. Efforts of large consortia (e.g. BLOODPAC) and societies (e.g. European Liquid Biopsy Society) have resulted in pre-analytical standards provided by the International Standards Organization (ISO). The ISO 20186-3:2019 is an international guideline for cfDNA examination, published in September 2019. The ISO guidelines are available upon payment of around €100 and provide requirements and recommendations on the handling, storage, processing and documentation of cfDNA examination from venous blood samples in a controlled environment. Despite this great initiative, wide-spread implementation of the ISO standard is limited. A recent preprint by Bonstingl *et al*. [34] evaluates the implementation of the standards in an actual research project. Despite the advantages of standardized workflows, one particularly important hurdle is the personnel required for managing and processing liquid biopsies. According to their study, it takes two full time biomedical scientists to fulfill the requirements of the ISO standards. Furthermore, as is also the case in our studies, blood samples for research purposes are often taken during clinical routine and forwarded to the lab. This makes it difficult to adhere to standards such temperature variations, tube inversions, identity of the person drawing blood… Of course, ISO standards should be implemented in a diagnostic setting where accredited tests are performed, but for research purposes, the most cumbersome recommendations are omitted.

Within CMG, we work with Standard Operation Procedures for our blood sample collections. We aim to use these SOPs for all our collections, to assure uniformity. As mentioned before, for research purposes the ISO standards are not fully implemented. However, we do aim to use the same blood collection tubes within studies,

the same centrifugation protocols, storage conditions and cfDNA extraction protocols. We have documents to report time of blood drawing, time of processing etc. Within one study, we adhere to one SOP. Nevertheless, our SOPs still might deviate between studies as research is always evolving, and unexpected changes may influence ongoing blood collections. Also, our SOPs might deviate from other research centers, which further complicates validation of biomarkers and assays. These are all very important aspects that will need evaluation in future studies.

FUTURE PERSPECTIVES

The era of liquid biopsies

Colorectal Cancer

Throughout the introduction and general discussion of this thesis, it has already become clear that much work has been performed on the subject of ctDNA methylation for CRC [35]. This has led to the first methylation blood-based test, the EpiProColon. However, as discussed, this test is still not optimal in view of sensitivity and specificity. It has been shown that combining more targets, increases the test sensitivity, which is important for early detection. This is exactly where IMPRESS could prove very useful. As discussed above, challenges must be overcome in the (near) future to be able to introduce a CRC-specific IMPRESS assay, but the fundaments for such assay are already provided in this thesis (**Chapter 7**).

Another liquid biopsy that has been gaining considerable attention in recent years, is urine. Whereas ctDNA has been extensively studied in blood, only few studies focus on the use of urine for CRC diagnosis [36,37]. The results of these studies indicate that total ctDNA from urine is a very promising diagnostic tool for future evaluation and implementation in the clinic. Here, IMPRESS might become an important new approach for urine-based ctDNA assays.

Studies for other liquid biopsy types are scarce. There is one study of Van 't Erve *et al.* [38] that describes the detection of ctDNA in peritoneal fluid. Furthermore, in other cancer types, breath condensates and saliva have been described as liquid biopsy. However, for CRC there are currently no studies on ctDNA in these fluid types. Within one of our ongoing studies, we have ethical approval for sampling breath condensates from CRC patients, so it might be a very interesting future project to analyze such samples for the presence of ctDNA (methylation).

Breast Cancer

Blood-based testing in BRCA has been less studied compared to CRC. cfDNA methylation analyses are mostly described in literature as discovery studies, with few methylation targets analyzed [39–41]. The most studied (hyper)methylated genes in BRCA include *BRCA1*, *RASSF1A* and *GSTP1* [40,41]. Once again, these studies show that combining targets increases assay sensitivity. The CCGA study from GRAIL included BRCA and has already been described in the discussion before. Specific for BRCA, TOO detection and specificity were high (93% and 99.3% respectively), but sensitivity for the validated cohort was only 30.5% across all stages [25,39]. Despite the low sensitivity, these results still highlight the potential of DNA methylation analysis of cfDNA for early BRCA detection. Besides DNA methylation, mutation detection of amongst others *PIK3CA* and *ESR1* in cfDNA

also gives promising results [41,42]. In the future, combining both mutation and methylation detection strategies might further enhance the sensitivity for early detection assays.

As opposed to CRC, there are no cfDNA methylation studies for BRCA in other fluids. A few studies describe the use of mutation detection in cfDNA extracted from urine [43–45]. Larger cohorts will be needed for validation of these findings towards clinical applications, but the use of urine for cancer detection includes important advantages in view of frequent and non-invasive testing. Furthermore, a few studies have investigated the use of nipple aspirate fluid (NAF) and saliva as liquid biopsies, but focused on miRNAs and proteomics instead of cfDNA [40,41].

Reflecting on this, early diagnosis whilst minimizing patient stress and discomfort is gradually becoming more feasible. The addition of IMPRESS to the DNA methylation detection research field, also contributes to the possibilities of developing affordable tests for cancer diagnosis. This can in turn lead to more options for screening in the population-based screening program. Of course, cost-effectiveness is a crucial part of an optimal screening program. Therefore, further research is necessary to determine the efficacy and economical value of alternative screening modalities, especially compared to current gold standard methods.

An important aspect of any novel assay or technology is its implementation in clinical settings. For IMPRESS, implementation in diagnostics is feasible, as sequencers are now routinely available in any diagnostic laboratory. IMPRESS does not require any other specific laboratory equipment, facilitating quick implementation in any lab. Another crucial step for clinical implementation is the use of IMPRESS in future clinical trials. For this, we have already established several research connections through previous clinical trials, with various hospitals in Europe. In this context, a therapy follow-up or minimal residual disease assay can be more easily implemented in the short term. Additionally, our screening assays for CRC and BRCA, which we are currently developing (**Chapter 7**), could complement current screening programs in the future to evaluate their accuracy and potentially replace them in the long term. Also, to ensure its valorization, we have already prioritized intellectual property by filing for a patent for the IMPRESS technology. Lastly, our biomarkers must be carefully evaluated in different populations to account for potential confounding effects of ethnicity. DNA methylation has been reported to vary among distinct populations, based on ancestry [46–48]. Furthermore, our previous studies have shown that this variation depends on the specific biomarker. For instance, Ibrahim *et al*. did not observe a confounding effect of ethnicity on *GSDME* methylation [49]. Therefore, the evaluation of future biomarker panels must be conducted in diverse populations to ensure smooth implementation.

All the efforts that are highlighted throughout this thesis prove that a shift towards the use of liquid biopsies in the clinic will take place in the future. There are a few important directions where liquid biopsies will become important medical tools. As previously mentioned, early detection through liquid biopsies is the most challenging problem that many researchers aim to tackle in the approaching years. Upcoming technological

advancements along with cost reductions will allow the screening of cancer via liquid biopsies. These will likely be blood samples at first, but will be expanded to urine and even other fluids in the long-term future. As single-cancer assays are easier to develop, these are expected to be implemented in a reasonable amount of time. However, once TOO prediction is more efficient and specific, it is more likely that multi-cancer screening tests will enter the market. Besides screening, liquid biopsies will also play an important role in monitoring of cancer such as MRD, therapy follow-up and treatment decisions. And lastly, an essential issue that might be solved using liquid biopsies, is the molecular subtyping of cancers. For both CRC and BRCA, but also for other cancer types, tumor heterogeneity affects treatment. By uncovering novel molecular markers such as DNA methylation and others and implementing analysis of these targets in liquid biopsy samples, personalized medicine for almost all patients will become more feasible.

The era of (multi) omics

In the past few years, novel omics strategies have gained attention for liquid biopsy applications in cancer. One of the most important examples is fragmentomics. This strategy is based upon multiple observations that cfDNA fragment length differs between healthy people and cancer patients. These differences include a more variable fragment size pattern in cancer patients, with a higher number of fragments below 150 bp and a lower number of fragments between 151-218 bp. In healthy persons, the average size is 167bp, whereas in cancer patients, this is 143 bp. Fragmentation is related to the TOO and leads to specific signatures of fragment size, nucleotide end motifs and break point motifs that can be researched [50–52]. In 2019, Cristiano *et al*. described the first test using fragmentomics with their technique called 'DELFI' (DNA evaluation of fragments for early interception). DELFI uses WGS to profile fragment size. In their study, 7 cancer types, including CRC and BRCA, were evaluated. A sensitivity ranging from 57% to 99% was reached at a specificity of 98%, with accurate TOO prediction of 75%. Upon combining with genomic strategies (mutations), 91% of cancers could be detected [23,50,53]. More recently, Bao *et al*. also described a fragmentomics-based MCED, where they use multiple cfDNA fragmentomic features stacked in a machine learning model. Even with only 1X coverage, they reached a sensitivity of 91.5% at 95% specificity, with 91.6% TOO prediction accuracy [54]. Fragmentomics is a very promising new field, and many studies provide proof of concept for limited sample sizes and limited cancer types. Larger validation studies are needed to prove its value in clinical applications.

As already extensively addressed in this thesis, DNA methylation plays different roles in cancer. The molecular alterations driving carcinogenesis can also affect noncoding RNAs and mRNA [55]. In this field of transcriptomics, most research has focused on microRNAs (miRNAs) which are small, non-protein coding RNAs that regulate gene expression. miRNA expression profiles have diagnostic value. Some miRNA levels are indeed altered in cancer, e.g. CRC, and regulate the *RAS* gene involved in tumorigenesis [56,57]. For example, several studies describe miR-92a and mi-R29a as potential plasma biomarkers for detection of colon adenomas and

carcinomas. Sensitivities of 73% and 83% respectively were reported at specificities around 80% [58]. Furthermore, long noncoding RNAs (lncRNAs) have also been reported to be useful for diagnostic purposes [57]. Gene expression analysis has already been implemented in clinical practice (e.g. Oncotype DX and MammaPrint), but only for prognostic purposes [57]. Only a short time ago, Stanley *et al*. published a diagnostic strategy based on single cell RNA-sequencing of cfDNA. Within their study, the classification models for CRC and BRCA could differentiate patients from matched controls with 84.7% and 90% accuracy respectively. This study once again highlights the future potential of transcriptomics for liquid biopsies [59]. Converting this towards our research, smMIPs have been described to be used for RNA sequencing by Arts *et al* [60]. Although outside of our current research scope, it could be very interesting to look into combining IMPRESS and smMIP-RNA-seq for increased sensitivity in the future.

Integrating multimodal information to improve assay performance has become more attractive and is one of the general future directions of the cancer detection research field. In view of this PhD thesis, combinations of epigenomics with other omics strategies is the most interesting approach. For single cancer assays, it has been reported that integrating multiple omics methods increases sensitivity compared to single modalities, if at least two different types of data are combined [61]. This is for example shown by Putcha *et al*. who combine genomics, epigenomics and proteomics through machine learning. Their combination of signals can detect early stage CRC with high sensitivity and specificity (92% and 90% respectively) [62]. Besides for single cancer assays, multi-omics are also increasingly used for MCED assay development. For one thing, studies show the potential of combining epigenomic and fragmentomic signals from cfDNA [23]. A recent example hereof is the GutSeer assay, commercially available through Singlera. It leverages fragment coverage, end-motif and cfDNA methylation features to detect and localize five deadly gastro-intestinal cancers. At specificities of 97%, this test reaches 86% sensitivity and 82% TOO accuracy [23,63]. Furthermore, Tomeva *et al*. describe the combination of cfDNA epigenomic (methylation), genomic (mutations) and transcriptomic (miRNA) analytes for increasing assay sensitivity. Via qPCR, they achieve high accuracy (95%) and very high sensitivity (97.9%) at specificity of 80% [64]. Looking at IMPRESS, we could easily combine genomic targets such as mutations and CNVs with epigenomic targets to improve diagnostic yield of our assays. Despite the recent research, it is too early to determine which combinations of molecular components will provide the highest diagnostic accuracy. Further research is needed to determine the optimal multi-omics strategy for clinical applications.

In the rapidly evolving multi-omics field, technologies that implement different omics, such long-read sequencing (epigenomics and genomics), will become the most important tools for clinical implementation. While long read sequencing is nowadays not feasible in a clinical setting due to its high cost, it is plausible that the cost will drop drastically in the coming years, allowing its implementation in a diagnostic context. Together with future innovations, it might become possible to create a 'molecular passport' for a patient, that

can be coupled to their medical file. Combining multi-omics with liquid biopsies will allow fast and personalized approaches for many patients.

Lastly, throughout all studies, it also becomes clear that computational methods are becoming of enormous importance in cancer research. In this view, artificial intelligence (AI) is rapidly growing and shaping the future. New applications are emerging, for example in pathology analysis, where AI-powered digital platforms assist pathologists in analyzing tissue samples more efficiently. In omics research, AI could become useful for analyzing large-scale data and identifying potential biomarkers. However, it will be crucial to carefully evaluate the use of such technology in future clinic applications.

CONCLUSION

Cancer remains a world-wide problem. Despite many efforts, non-invasive, early detection is still difficult. Addressing this challenge relies upon the refinement of robust biomarkers and innovative diagnostic methodologies. This thesis explored the examination of aberrant DNA methylation profiling in cancer and the consequent development of diagnostic biomarkers and assays. Through comprehensive analysis, we have investigated novel biomarkers to augment cancer detection, using both conventional and state-of-the-art technologies. We contributed to the increasing interest in early cancer detection of multiple cancers, while also paving the way for future applications in CRC and BRCA detection. Our assays hold promise as non-invasive tools for early cancer diagnosis. While concluding this Ph.D. study, further research is crucial to translate findings into clinical practice. Looking ahead, insights from this work offer potential for developing liquid biopsy assays and multi-omics strategies in cancer research and beyond.

References

1. Oh CK, Cho Y-S. Pathogenesis and biomarkers of colorectal cancer by epigenetic alteration. Intest Res. 2024;

2. Liu C, Tang H, Hu N, Li T. Methylomics and cancer: the current state of methylation profiling and marker development for clinical care. Cancer Cell Int. BioMed Central Ltd; 2023.

3. Müller D, Győrffy B. DNA methylation-based diagnostic, prognostic, and predictive biomarkers in colorectal cancer. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2022;1877:188722.

4. Bonetti G, Madeo G, Michelini S, Ricci M, Cestari M, Gadler M, et al. Omics sciences and precision medicine in breast and ovarian cancer. Clin Ter. 2023;174:104–18.

5. Laugsand EA, Brenne SS, Skorpen F. DNA methylation markers detected in blood, stool, urine, and tissue in colorectal cancer: a systematic review of paired samples. Int J Colorectal Dis. 2021;36:239–51.

6. Dong L, Wang S, Fu B, Wang J. Evaluation of droplet digital PCR and next generation sequencing for characterizing DNA reference material for KRAS mutation detection. Sci Rep. 2018;8:9650.

7. Singh RR. Next-Generation Sequencing in High-Sensitive Detection of Mutations in Tumors: Challenges, Advances, and Applications. Journal of Molecular Diagnostics. Elsevier B.V.; 2020. p. 994–1007.

8. Ibrahim J, Op de Beeck K, Fransen E, Peeters M, Van Camp G. Genome-wide DNA methylation profiling and identification of potential pan-cancer and tumor-specific biomarkers. Mol Oncol. 2022;16:2432–47.

9. Marcellinaro R, Spoletini D, Grieco M, Avella P, Cappuccio M, Troiano R, et al. Colorectal Cancer: Current Updates and Future Perspectives. J Clin Med. 2023;13:40.

10. Schwartzberg L, Broder MS, Ailawadhi S, Beltran H, Blakely LJ, Budd GT, et al. Impact of early detection on cancer curability: A modified Delphi panel study. PLoS One. 2022;17.

11. Mittra I, Mishra GA, Dikshit RP, Gupta S, Kulkarni VY, Shaikh HKA, et al. Effect of screening by clinical breast examination on breast cancer incidence and mortality after 20 years: prospective, cluster randomised controlled trial in Mumbai. BMJ. 2021;n256.

12. Ma Z qiang, Richardson LC. Cancer Screening Prevalence and Associated Factors Among US Adults. Prev Chronic Dis. 2022;19:1–5.

13. NHS. Signs and symptoms of Cancer. https://www.nhs.uk/conditions/cancer/symptoms/.

14. Brito-Rocha T, Constâncio V, Henrique R, Jerónimo C. Shifting the Cancer Screening Paradigm: The Rising Potential of Blood-Based Multi-Cancer Early Detection Tests. Cells. MDPI; 2023.

15. Thien Chi Nguyen V, Hieu Nguyen T, Nhat Tan Doan N, Mong Quynh Pham T, Thi Huong Nguyen G, Dat Nguyen T, et al. Viet Nam; 5 National Cancer Hospital. Thu Duc City Hospital [Internet]. [cited 2024 Feb 1];6. Available from: https://doi.org/10.7554/eLife.89083https://doi.org/10.7554/eLife.89083

16. Jamshidi A, Liu MC, Klein EA, Venn O, Hubbell E, Beausang JF, et al. Evaluation of cell-free DNA approaches for multi-cancer early detection. Cancer Cell. 2022;40:1537-1549.e12.

17. Hackshaw A, Clarke CA, Hartman A-R. New genomic technologies for multi-cancer early detection: Rethinking the scope of cancer screening. Cancer Cell. 2022;40:109–13.

18. Hathaway C, Paetsch P, Li Y, Wu J, Asgarian S, Parker A, et al. Association of Breast Cancer Screening Behaviors With Stage at Breast Cancer Diagnosis and Potential for Additive Multi-Cancer Detection via Liquid Biopsy Screening: A Claims-Based Study. Front Oncol. 2021;11.

19. Tafazzoli A, Ramsey SD, Shaul A, Chavan A, Ye W, Kansal AR, et al. The Potential Value-Based Price of a Multi-Cancer Early Detection Genomic Blood Test to Complement Current Single Cancer Screening in the USA. Pharmacoeconomics [Internet]. 2022 [cited 2024 Feb 12];40:1107. Available from: /pmc/articles/PMC9550746/

20. Myers RE, Hallman MH, Shimada A, DiCarlo M, Davis K, Leach WT, et al. Primary Care Patient Interest in Multi-Cancer Early Detection for Cancer Screening. J Pers Med. 2023;13.

21. Gelhorn H, Ross MM, Kansal AR, Fung ET, Michael ·, Seiden V, et al. Patient Preferences for Multi-Cancer Early Detection (MCED) Screening Tests. 2023 [cited 2024 Feb 12];16:43–56. Available from: https://doi.org/10.1007/s40271-022-00589-5

22. LeeVan E, Pinsky P. Predictive Performance of Cell-Free Nucleic Acid-Based Multi-Cancer Early Detection Tests: A Systematic Review. Clin Chem [Internet]. 2024 [cited 2024 Feb 1];70:90–101. Available from: https://dx.doi.org/10.1093/clinchem/hvad134

23. Zhang K, Fu R, Liu R, Su Z. Circulating cell-free DNA-based multi-cancer early detection. Trends Cancer. 2024;10:161–74.

24. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, Liu MC, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Annals of Oncology. 2020;31:745–59.

25. Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Annals of Oncology. 2021;32:1167–77.

26. Liu MC. Transforming the landscape of early cancer detection using blood tests—Commentary on current methodologies and future prospects. Br J Cancer. 2021;124:1475–7.

27. Neal RD, Johnson P, Clarke CA, Hamilton SA, Zhang N, Kumar H, et al. Cell-Free DNA–Based Multi-Cancer Early Detection Test in an Asymptomatic Screening Population (NHS-Galleri): Design of a Pragmatic, Prospective Randomised Controlled Trial. Cancers (Basel) [Internet]. 2022 [cited 2024 Feb 12];14:4818. Available from: /pmc/articles/PMC9564213/

28. Farooq M, Leevan E, Ahmed J, Ko B, Shin S, De Souza A, et al. Blood-based multi-cancer detection: A state-of-the-art update. 2024 [cited 2024 Feb 1]; Available from: https://doi.org/10.1016/j.currproblcancer.2023.101059

29. Konigsberg IR, Barnes B, Campbell M, Davidson E, Zhen Y, Pallisard O, et al. Host methylation predicts SARS-CoV-2 infection and clinical outcome. Communications Medicine. 2021;1:42.

30. Pérez RF, Alba-Linares JJ, Tejedor JR, Fernández AF, Calero M, Román-Domínguez A, et al. Blood DNA Methylation Patterns in Older Adults With Evolving Dementia. The Journals of Gerontology: Series A. 2022;77:1743–9.

31. Kerminen H, Marzetti E, D'Angelo E. Biological and Physical Performance Markers for Early Detection of Cognitive Impairment in Older Adults. J Clin Med. 2024;13:806.

32. Chiapperino L, Paneni F. Why epigenetics is (not) a biosocial science and why that matters. Clin Epigenetics. 2022;14:144.

33. Downs BM, Sukumar S. Capturing ctDNA from Unaltered Stationary and Flowing Plasma with dCas9. ACS Appl Mater Interfaces. 2022;14:24113–21.

34. Bonstingl L, Skofler C, Ulz C, Zinnegger M, Sallinger K, Schönberger J, et al. Clinical application of ISO and CEN/TS standards for liquid biopsies - information everybody wants but nobody wants to pay for. Available from: https://doi.org/10.1101/2023.12.04.23299422

35. Müller D, Győrffy B. DNA methylation-based diagnostic, prognostic, and predictive biomarkers in colorectal cancer. Biochim Biophys Acta Rev Cancer. Elsevier B.V.; 2022.

36. Song T, Mao F, Shi L, Xu X, Wu Z, Zhou J, et al. Urinary measurement of circulating tumor DNA for treatment monitoring and prognosis of metastatic colorectal cancer patients. Clin Chem Lab Med [Internet]. 2019 [cited

2024 Apr 11];57:268–75. Available from: https://www.degruyter.com/document/doi/10.1515/cclm-2017-0675/html

37. Bach S, Paulis I, Sluiter NR, Tibbesma M, Martin I, van de Wiel MA, et al. Detection of colorectal cancer in urine using DNA methylation analysis. Sci Rep. 2021;11.

38. van't Erve I, Rovers KP, Constantinides A, Bolhuis K, Wassenaar EC, Lurvink RJ, et al. Detection of tumor-derived cell-free <scp>DNA</scp> from colorectal cancer peritoneal metastases in plasma and peritoneal fluid. J Pathol Clin Res. 2021;7:203–8.

39. Keup C, Kimmig R, Kasimir-Bauer S. The Diversity of Liquid Biopsies and Their Potential in Breast Cancer Management. Cancers (Basel) [Internet]. 2023 [cited 2024 Feb 1];15. Available from: /pmc/articles/PMC10670968/

40. Li J, Guan X, Fan Z, Ching L-M, Li Y, Wang X, et al. Non-Invasive Biomarkers for Early Detection of Breast Cancer. Cancers (Basel). 2020;12:2767.

41. Delmonico L, Alves G, Bines J. Cell free DNA biology and its involvement in breast carcinogenesis. Adv Clin Chem. Academic Press Inc.; 2020. p. 171–223.

42. Beaver JA, Jelovac D, Balukrishna S, Cochran RL, Croessmann S, Zabransky DJ, et al. Detection of Cancer DNA in Plasma of Patients with Early-Stage Breast Cancer. Clinical Cancer Research. 2014;20:2643–50.

43. Zuo Z, Tang J, Cai X, Ke F, Shi Z. Probing of breast cancer using a combination of plasma and urinary circulating cell-free DNA. Biosci Rep. 2020;40.

44. Guan G, Wang Y, Sun Q, Wang L, Xie F, Yan J, et al. Utility of urinary ctDNA to monitoring minimal residual disease in early breast cancer patients. Cancer Biomarkers. 2020;28:111–9.

45. Shuai Y, Ma Z, Ju J, Wei T, Gao S, Kang Y, et al. Liquid-based biomarkers in breast cancer: looking beyond the blood. J Transl Med [Internet]. 2023 [cited 2024 Mar 12];21:809. Available from: /pmc/articles/PMC10644618/

46. Lerner L, Winn R, Hulbert A. Lung cancer early detection and health disparities: the intersection of epigenetics and ethnicity. J Thorac Dis. 2018;10:2498–507.

47. Chan MH, Merrill SM, Konwar C, Kobor MS. An integrative framework and recommendations for the study of DNA methylation in the context of race and ethnicity. Discover Social Science and Health. 2023;3:9.

48. Telonis AG, Rodriguez DA, Spanheimer PM, Figueroa ME, Goel N. Genetic Ancestry-specific Molecular and Survival Differences in Admixed Patients With Breast Cancer. Ann Surg. 2024;279:866–73.

49. Ibrahim J, Op de Beeck K, Fransen E, Peeters M, Van Camp G. The Gasdermin E Gene Has Potential as a Pan-Cancer Biomarker, While Discriminating between Different Tumor Types. Cancers (Basel). 2019;11:1810.

50. Thierry AR. Circulating DNA fragmentomics and cancer screening. Cell Genomics [Internet]. 2023 [cited 2024 Mar 4];3:100242. Available from: /pmc/articles/PMC9903826/

51. Qi T, Pan M, Shi H, Wang L, Bai Y, Ge Q. Cell-Free DNA Fragmentomics: The Novel Promising Biomarker. Int J Mol Sci [Internet]. 2023 [cited 2024 Mar 4];24. Available from: /pmc/articles/PMC9866579/

52. Lo YMD, Han DSC, Jiang P, Chiu RWK. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. Science (1979) [Internet]. 2021 [cited 2024 Mar 4];372. Available from: https://www.science.org/doi/10.1126/science.aaw3616

53. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. Nature. 2019;570:385–9.

54. Bao H, Wang Z, Ma X, Guo W, Zhang X, Tang W, et al. Letter to the Editor: An ultra-sensitive assay using cell-free DNA fragmentomics for multi-cancer early detection Open Access †. Mol Cancer [Internet]. 2021 [cited

2024 Mar 4];21:129. Available from: http://creativecommons.org/licenses/by/4.0/.TheCreativeCommonsPublicDomainDedicationwaiver

55. Bustin SA, Murphy J. RNA biomarkers in colorectal cancer. Methods [Internet]. 2013 [cited 2024 Apr 11];59:116–25. Available from: https://pubmed.ncbi.nlm.nih.gov/23079397/

56. Meltzer PS. Cancer genomics: small RNAs with big impacts. Nature [Internet]. 2005 [cited 2024 Apr 11];435:745–6. Available from: https://pubmed.ncbi.nlm.nih.gov/15944682/

57. Tsimberidou AM, Fountzilas E, Bleris L, Kurzrock R. Transcriptomics and solid tumors: The next frontier in precision cancer medicine. Semin Cancer Biol [Internet]. 2022 [cited 2024 Feb 28];84:50–9. Available from: http://creativecommons.org/licenses/by/4.0/

58. Gallardo-Gómez M, De Chiara L, Álvarez-Chaver P, Cubiella J. Colorectal cancer screening and diagnosis: omics-based technologies for development of a non-invasive blood-based method. Expert Rev Anticancer Ther. Taylor and Francis Ltd.; 2021. p. 723–38.

59. Stanley KE, Jatsenko T, Tuveri S, Sudhakaran D, Lannoo L, Van Calsteren K, et al. Cell type signatures in cell-free DNA fragmentation profiles reveal disease biology. Nat Commun. 2024;15.

60. Arts P, Van Der Raadt J, Van Gestel SHC, Steehouwer M, Shendure J, Hoischen A, et al. Quantification of differential gene expression by multiplexed targeted resequencing of cDNA. Nat Commun [Internet]. 2017 [cited 2024 Apr 11];8. Available from: /pmc/articles/PMC5424154/

61. Chen G, Zhang J, Fu Q, Taly V, Tan F. Integrative analysis of multi-omics data for liquid biopsy. 2023 [cited 2024 Mar 4]; Available from: https://doi.org/10.1038/s41416-022-02048-2

62. Putcha G, Liu T-Y, Ariazi E, Bertin M, Drake A, Dzamba M, et al. Blood-based detection of early-stage colorectal cancer using multiomics and machine learning. Journal of Clinical Oncology. 2020;38:66–66.

63. Yang X, He D, Xiong Z, Yan B, Li Q, Guo D, et al. Early detection and localization of major gastrointestinal cancers by combining DNA methylation and fragmentomic signatures generated from a single cfDNA targeted sequencing assay. Journal of Clinical Oncology. 2023;41:330–330.

64. Tomeva E, Switzeny OJ, Heitzinger C, Hippe B, Haslberger AG. Comprehensive Approach to Distinguish Patients with Solid Tumors from Healthy Controls by Combining Androgen Receptor Mutation p.H875Y with Cell-Free DNA Methylation and Circulating miRNAs. Cancers (Basel). 2022;14:462.

List of Abbreviations

LIST OF ABBREVIATIONS

%CV	Intra-class coefficient of variability
(q)MSP	(quantitative) Methylation-Specific PCR
5caC	5' carboxyl-cytosine
5fC	5'formyl-cytosine
5hmC	5' hydroxymethylated cytosine
5mC	5' methylated cytosine
A-C-N	Adenoma-Carcinoma-Normal
ATCC	American Type Culture Collection
AUC	Area under the curve
BRCA	Breast cancer
BSAS	Bisulfite Amplicon Sequencing
BSPP	Bisulfite padlock probes
CA19-9	Carbohydrate antigen 19-9
CCGA	Circulating cell free genome atlas
CEA	Carcinoembrionic antigen
CE-IVD	European certified in vitro diagnostics
cfDNA	Cell free DNA
CGI	CpG Island
CHARM	Comprehensive High-throughput Arrays for Relative Methylation
CIMP	CpG island methylator phenotype
CNV	Copy number variation
COBRA	Combined Bisulfite Restriction Analysis
COLIPAN	Collection of liquid biopsies in pancancer patients
COSMIC	Catalogue of Somatic Mutations in Cancer
CRC	Colorectal cancer
CUP	Cancer of Unknown Primary
cvAUC	Cross-validated area under the curve
ddPCR	Droplet digital PCR
DE DMPs	Double evidence" differentially methylated probes
DHU	Dihydrouracil
Di	Depth of invasion
DMB	Differentially Methylated Blocks
DMP	Differentially Methylated Probes
DMR	Differentially Methylated Regions
DREAM	Digital Restriction Enzyme Analysis of Methylation
DSZM	German Collection of Microorganism and Cell cultures
ELISA	Enzyme-Linked Immune-Sorbent Assay
ELSA-Seq	Enhanced Linear-Splinter Amplification Sequencing
EM-seq	Enzymatic Methyl sequencing
EMT	Epithelial to mesenchymal transition

EMX1	Empty Spiracles-Like Protein 1
EpiGScar	Epigenomics and Genomics of Single cells analyzed by restriction
ER	Estrogen receptor
ESMO	European Society of Medical Oncology
ETD	Extra tumoral Deposit
FDR	False Discovery Rate
FFPE	Formalin-fixed paraffin-embedded
FIT	Fecal Immunohistochemic Test
FRET	Förster Resonance Energy Transfer
FS	Flexible sigmoidoscopy
GEO	Gene Expression Omnibus
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
Hammer-Seq	Hairpin-Assisted Mapping of Methylation of Replicated DNA Sequencing
HCC	Hepatocellular carcinoma
HELP	HpaII-tiny fragment Enrichment by Ligation-mediated PCR
HR	Hormone Receptor
IBD	Inflammatory bowel disease
IC	Interval Cancer
ICGC	International Cancer Genome Consortium
IHC	Immunohistochemistry
IMPRESS	Improved Methylation Profiling using Restriction Enzymes and smMIP Sequencing
ISO	International standards organisation
LDA	Linear discriminant analysis
LDT	Laboratory developed test
LNM	Lymph node metastasis
LOB	Limit of blank
LOD	Limit of detection
LUMA	Luminometric Methylation Assay
LVI	Lymphovascular invasion
MBD	Methyl-CpG binding domain
MCD	Multi-cancer detection
MCED	Multi-cancer early detection
MCTA-Seq	Methylated CpG Tandem Amplification and Sequencing
MDS	Multidimensional scaling
MeDIP	Methylated DNA immunoprecipitation
MED-seq	Methylated DNA sequencing
MHC	Major histocompatibility complex
MIMIC	Minimal methylation classifier
MMR	Mismatch repair
MRD	Minimal residual disease
MSI	Microsattelite instability
MS-MLPA	Methylation-Specific Multiplex Ligation Probe Amplification
MSRE	Methylation-sensitive restriction enzymes

MT-sDNA	Multitarget stool DNA
Mwcnt	Multi-Walled Carbon Nano Tubes
NAF	Nipple aspirate fluid
NCI	National cancer institute
NET	Neuroendocrine tumor
NFQ-MGB	Non-fluorescent quencher minor groove binder
NGS	Next-generation Sequencing
NMBIC	Non-muscle invasive bladder cancer
NST	No specific Type
NXPH1	Neurexophilin 1
OBBPA-ddPCR	Optimized bias-based pre-amplification ddPCR
ONT	Oxford Nanopore Technologies
OR	Odds Ratio
PBAT	Post bisulfite adaptor tagging
PR	Progesteron receptor
ROC	Receiver operating characteristic (curve)
ROI	Region of Interest
RRBS	Reduced Representation Bisulfite Sequencing
SDC2	Syndecan 2
SD	Screen Detected
SERS	Surface-Enhanced Raman Spectrophotometry
smMIP	Single molecule Molecular Inversion Probes
SMRT	Single Molecule Real-Time
TAPS	TET-Assisted Pyridine borane Sequencing
TAT	Turn around time
TCGA	The Cancer Genome Atlas
TcP	Tumor cell percentage
TET	Ten-eleven translocation
TGS	Third Generation Sequencing
TNBC	Triple negative breast cancer
TNM	Tumor, Node, Metastasis
TOO	Tissue of origin
t-SNE	T-distributed stochastic neighbor embedding
TSS	Transcription start site
UMI	Unique molecular identifier
UTR	Untranslated region
WGA	Whole genome amplified
WGBS	Whole Genome Bisulfite Sequencing

Curriculum Vitae

CURRICULUM VITAE

PERSONAL DETAILS

Name	Isabelle Neefs
Address	Pijpelstraat 44, 2221 Booischot
Phone number	+32 478 17 99 29
E-mail	Isabelle.neefs@uantwerpen.be
	Isabelle_neefs@hotmail.com
Date of Birth	26/11/1996
Place of Birth	2610 Wilrijk

EDUCATION AND ACADEMIC CAREER

2019 – Present

PhD in Biomedical Sciences, FWO Strategic Basic Fellow

Centre of Medical Genetics and Centre for Oncological Research, University of Antwerp

Title: Enhancing cancer detection through novel DNA methylation strategies and biomarkers

Promoters: Prof. Dr. Guy Van Camp, Prof. Dr. Ken Op de Beeck, Prof. Dr. Marc Peeters

Supervisor: Prof. Dr. Wiebren Tjalma

2017 – 2019

Master in Biomedical Sciences

University of Antwerp

Major: Molecular and Cellular Biomedical Sciences

Minor: Research

Graduated *Summa Cum Laude*

Master thesis

Translational Research and Microfluidics Group, Université Paris Descartes, FRANCE (Erasmus exchange program)

Title: *Methylated circulating tumor DNA as a biomarker for detection of advanced-stage ovarian cancer*

Promoters: Dr. Valérie Taly, Prof. Dr. Guy Van Camp

Co-promoters: Dr. Shufang-Wang Renault, Prof. Dr. Ken Op de Beeck

2014 – 2017 **Bachelor in Biomedical Sciences**

University of Antwerp

Graduated *Magna Cum Laude*

Bachelor thesis

Radiology, Antwerp University Hospital

Title: *Influence of pelvic anatomy on the development and progression of adolescent idiopathic scoliosis*

Promoter: Prof. Dr. Johan Van Goethem

SCIENTIFIC WORK

PUBLICATIONS

Neefs I, Tran T, Ferrari A, Janssens S, Van Herck K, Fransen E, Op de Beeck K, Van Camp G, Peeters M, Hoeck S, Van Hal G (2024). **Clinicopathological and molecular differences between stage IV screen-detected and interval colorectal cancers in the Flemish screening program.** Manuscript submitted

Neefs I, Ibrahim J, Peeters M, Van Camp G, Op de Beeck K (2024). **From Classic To Recent DNA Methylation Detection Methods in cancer: The current technology Landscape.** Manuscript submitted

Neefs I, De Meulenaere N, Vanpoucke T, Vandenhoeck J, Peeters D, Peeters M, Van Camp G, Op de Beeck K (2024). **Simultaneous detection of eight cancer types using a multiplex droplet digital PCR assay.** Revisions submitted at Molecular Oncology

Vandenhoeck J*, Neefs I*, Vanpoucke T*, Ibrahim J, Suls A, Peeters D, Schepers A, Hoischen A, Fransen E, Peeters M, Van Camp G, Op de Beeck K (2024). **IMPRESS: Improved Methylation Profiling using Restriction Enzymes and smMIP sequencing, Combined with a New Biomarker Panel, Creating a Multi-Cancer Detection Assay.** Revisions submitted at British Journal of Cancer

Neefs I (2024). **Vroegtijdig kanker ontmaskeren via een bloedstaal.** Eos Blogs

Neefs I, Vandenhoeck J, Vanpoucke T, Ibrahim J, Suls A, Hoischen A, Peeters M, Op de Beeck K, Van Camp G (2023). **IMPRESS: Improved Multiplex Methylation Profiling Using Restriction Enzymes and smMIP Sequencing for Highly Sensitive Multi-Cancer Detection.** The Journal of Liquid Biopsy, doi:10.1016/j.jlb.2023.100073

Janssens K*, Neefs I*, Ibrahim J*, Schepers A, Pauwels P, Peeters M, Van Camp G, Op de Beeck K (2023). **Epigenome-wide methylation analysis of colorectal carcinoma, adenoma and normal tissue reveals**

novel biomarkers addressing unmet clinical needs. Clinical Epigenetics, doi:10.1186/s13148-023-01516-7

Neefs I, Vandenhoeck J, Vanpoucke T, Ibrahim J, Suls A, Hoischen A, Peeters M, Op de Beeck K, Van Camp G (2023). **A novel bisulfite free, highly multiplexable assay to detect DNA methylation levels.** Molecular Oncology 17S1, doi:10.1002/1878-0261.13471

Neefs I, Tjalma W, Ibrahim J, Croes L, Peeters M, Van Camp G, Op de Beeck K (2023). **Early detection of breast cancer in liquid biopsies using DNA methylation markers.** The Breast 62S1, S15–S136:PO38: s30-s31, doi: 10.1016/S0960-9776(23)00157-1

Ferrari A, Neefs I, Hoeck S, Peeters M, Van Hal G (2021). **Op weg naar nieuwe niet-invasieve methodes voor dikkedarmkanker-screening: een uitgebreide review.** Oncohemato, vol 15 (6), p42 - 46

Ferrari A, Neefs I, Hoeck S, Peeters M, Van Hal G (2021). **Towards Novel Non-Invasive Colorectal Cancer Screening Methods: A Comprehensive Review.** Cancers, doi:10.3390/cancers13081820

PATENTS

Van Camp G, Peeters M, Op de Beeck K, Suls A, Neefs I, Vandenhoeck J (Inventors). **METHYLATION DETECTION ASSAY (patent)**. European EP4211265A1, International WO2022053637A1, filed in September 2021

ORAL PRESENTATIONS

Neefs I., De Meulenaere N., Vanpoucke T., Vandenhoeck J., Peeters D., Peeters M., Van Camp G., Op de Beeck K. – **Simultaneous detection of eight cancer types using a multiplex droplet digital PCR assay** at the European Digital PCR Symposium 2024 in Gent, BELGIUM, February 2024

Neefs I.*, Vandenhoeck J.*, Vanpoucke T.*, Ibrahim J., Peeters D., Suls A., Hoischen A., Peeters M., Van Camp G., Op de Beeck K. – **IMPRESS: Improved Methylation Profiling using Restriction Enzymes and smMIP sequencing, Combined with a New Biomarker Panel, Creating a Multi-Cancer Detection Assay** at the Cancer Research Day Uantwerp in Antwerp, BELGIUM, October 2023

Neefs I.*, Vandenhoeck J.*, Vanpoucke T.*, Ibrahim J., Suls A., Hoischen A., Peeters M., Van Camp G., Op de Beeck K. – **A novel bisulphite free, highly-multiplexeable assay to detect DNA methylation levels** at the 5th cfDNA day in Houten, THE NETHERLANDS, June 2023

Neefs I., Peeters M., Van Camp G, Op de beeck K. – **Development of novel (epi)genetic technologies for sensitive DNA detection** at the Genomed Scientific Advisory Board Meeting, Antwerp, BELGIUM, March 2022

POSTER PRESENTATIONS

Neefs I., De Meulenaere N., Vanpoucke T., Vandenhoeck J., Peeters D., Peeters M., Van Camp G., Op de Beeck K. – **A novel multiplex droplet digital PCR assay for simultaneous detection of eight frequent cancer types**

at the 24th annual meeting of the Belgian Society for Human Genetics (BeSHG), Leuven, BELGIUM, April 2024

Neefs I., Vandenhoeck J., Vanpoucke T., Ibrahim J., Suls A., Hoischen A., Peeters M., Van Camp G., Op de Beeck K. – **IMPRESS: Improved methylation profiling using restriction enzymes and smMIP sequencing for highly sensitive multi-cancer detection** at the 5th annual congress of the International Society of Liquid Biopsy (ISLB), Madrid, SPAIN, November 2023

Neefs I., Vandenhoeck J., Vanpoucke T., Ibrahim J., Suls A., Hoischen A., Peeters M., Van Camp G., Op de Beeck K. – **IMPRESS: a novel bisulfite free, highly-multiplexable assay to detect DNA methylation levels** at 24th annual congress of the European Association for Cancer Research (EACR), Turin, ITALY, June 2023

Neefs I., Tjalma W., Ibrahim J., Croes L., Peeters M., Van Camp G., Op de Beeck K. – **Early detection of breast cancer in liquid biopsies using DNA methylation markers** at the 18th St. Gallen International Breast Cancer Conference in St. Gallen, SWITZERLAND, March 2023

Neefs I.*, Janssens K.*, Ibrahim J.*, Pauwels P., Schepers A., Peeters M., Van Camp G., Op de Beeck K. - **Epigenome-wide methylation analysis of colorectal carcinoma, adenoma and normal tissue reveals novel biomarkers for potential clinical use** at the Cancer Research Day in Antwerp, BELGIUM, October 2022

Neefs I.*, Janssens K.*, Ibrahim J., Schepers A., Peeters M., Van Camp G., Op de Beeck K. – **Methylation differences between colorectal normal, adenoma and carcinoma tissue have great potential to become important new biomarkers in the clinic** at the Cancer Genomics conference from the European Association for Cancer Research (EACR) in Oxford, UNITED KINGDOM, June 2022

GRANTS AND AWARDS

2022	**2-year PhD Fellowship Strategic Basic Research** Research Foundation Flanders
2020	**2-year PhD Fellowship Strategic Basic Research** Research Foundation Flanders
2019	**Prize for the Most Meritorious Master Thesis of the year** Department of Biomedical Sciences, University of Antwerp

ADDITIONAL TRAINING AND COURSES

2023	**Science communication writing course** SciComm Academy together with FWO, Language Union and Eos Science

2023	**Inside a company** Antwerp Doctoral School, University of Antwerp
2022	**IHC-GCP E6 (R2)** The Global Health Network
2022	**Grow to Lead** Antwerp Doctoral School, University of Antwerp
2022	**NGS course** University of Antwerp
2022	**Personal Effectiveness** Antwerp Doctoral School, University of Antwerp
2020	**Giving Presentations in English** Antwerp Doctoral School, University of Antwerp
2019	**FELASA Experiment leader, category C** University of Antwerp
2018	**Academic Writing in English: Writing Papers and Dissertations** Linguapolis, University of Antwerp

EDUCATIONAL ACTIVITIES

MASTER STUDENTS

2020 – 2021	**Laura Mariën, Master in Biomedical Sciences** University of Antwerp Title: *Developing a novel methylation detection assay for early pancancer detection in liquid biopsies*
2019 – 2020	**Ana Regina De Abreu, Master in Biomedical Sciences** University of Antwerp Title: *The position of circulating tumor DNA in clinical management of colorectal cancer*

BACHELOR STUDENTS

2023	**Jitske Driessen, Bachelor in Biochemistry (Honours internship)** University of Antwerp Title: *Verificatie van in silico kankerbiomerkers*
2020	**Quinten Van Looy, Bachelor in Biochemistry** University of Antwerp Title: *Differentiële methylatie op CpGs tussen gezonde en tumorcellen*

Acknowledgements | Dankwoord

DANKWOORD - ACKNOWLEDGEMENTS

"*Let us be grateful to the people who make us happy; they are the charming gardeners who make our souls blossom*"

Marcel Proust

DANKWOORD

Ongelofelijk, na bijna 5 jaar zit mijn doctoraat erop. Het was een traject vol bloed (letterlijk – 10 liter om precies te zijn), zweet en tranen (van miserie maar ook van geluk). Dit hele avontuur is reeds gestart vooraleer ik aan mijn doctoraat begon. Al vanaf de allereerste les genetica van Guy en oncologie van Ken was ik geboeid door dit onderwerp en bij het infomoment voor de masterthesis was ik er dan ook van overtuigd dat ik in hun lab wou werken. Onze eerste meeting gaf mij direct een goed gevoel en dankzij hen beide kon ik terecht in het lab van Valérie, waar ik mijn thesis mocht uitvoeren. Daar werd mijn passie voor biomerkers gevoed en ik was dan ook enorm blij wanneer ik de kans kreeg om mijn doctoraat te starten op 24 september 2019. Tot op de dag van vandaag heb ik vanuit alle hoeken steun en hulp gekregen om dit hele project tot een goed einde te brengen en ik zou hiervoor graag alle mensen willen bedanken die hun steentje hebben bijgedragen.

Eerst en vooral zou ik graag mijn promotoren Guy, Ken en prof. Peeters willen bedanken. Zij hebben mij niet alleen de kans gegeven om aan dit doctoraat te starten, maar ook doorheen mijn doctoraat hebben ze mij heel veel kansen gegeven om bijvoorbeeld op congres te gaan, om mijn ideeën uit te testen in het labo, om cursussen te volgen… Guy en Ken, bedankt voor jullie steun, jullie begrip als het eens wat moeilijker ging, jullie inzichten wanneer ik vast zat en jullie geduldige begeleiding doorheen het hele traject. Bedankt om in mij te geloven vanaf het prille begin en voor alle relativerende woorden. Bedankt om mij moed in te spreken en rust te bieden wanneer we een deadline naderden. Ik heb van jullie beiden veel geleerd, zowel wetenschappelijke alsook mensenkennis.

Bedankt ook prof. Peeters, om altijd kritisch te zijn, om de juiste vragen te stellen en om altijd tijd te maken, ook wanneer u eigenlijk geen tijd had. Ik weet dat dit niet evident was voor u als arts – laat staan als CEO, maar ik apprecieer alle moeite enorm.

Daarnaast wil ik ook heel graag Prof. Tjalma bedanken om als begeleider op te treden in mijn onderzoek. Uw nauwe opvolging, zeker bij de borstkanker staalcollectie, heeft mij deugd gedaan. Bedankt om mij kansen aan te reiken en om steeds in mij te geloven. Dankzij uw kritische blik, die telkens gepaard ging met motiverende woorden, bleef ik er zelf ook in geloven. Ik zou graag ook alle leden van mijn jury willen bedanken., Prof. Dr. Wim Vanden Berghe, Prof. Dr. Christophe Deben, Prof. Katleen De Preter and Dr. Charlotte D'Hulst, bedankt om de tijd te nemen om mijn thesis te lezen en voor de boeiende discussies omtrent mijn werk. Extra bedankt aan Christophe voor alle hulp met de robot en om zo Chapter 7 mee een feit te maken!

Oneindig veel dank gaat ook naar alle patiënten die deelgenomen hebben in de studies die ik opgestart heb tijdens mijn PhD project. Dankzij hun onbaatzuchtige deelname hebben we een zeer mooie collectie cfDNA aangelegd die voor mijn toekomstige opvolger een zeer nuttige validatie set zou kunnen betekenen.

Dit brengt mij naadloos over naar een heel aantal personen op het UZA die ik heel dankbaar ben voor hun medewerking. Eerst en vooral Dr. Heiko De Schepper, Martine en Frauke, dankzij wie mijn allereerste studie ooit supervlot verlopen is. Heel erg bedankt voor jullie tips, inzet en motivatie. Ook alle verpleegkundigen van de dienst Gastro-enterologie, bedankt om mee te werken aan deze studie ondanks dat het voor jullie geen makkelijke periode was met COVID-19. Ook dankzij jullie ging het verzamelen van een leien dakje. Om even binnen de dienst te blijven, bedank ik ook heel graag Dr. Pieter Dewint, Dr. Souad Bouhadan, Sophie en Eveline voor hun enthousiasme en samenwerking voor de EMR-studie. Bedankt om jullie tijd hieraan te willen spenderen, dit heeft wederom gezorgd voor een mooie collectie aan zowel bloed als weefselstalen. Ook Prof. Dr. Niels Komen, Dr. Sylvie Vandenbroeck en alle studieco's (Dorien x2 en Lieselotte) ben ik dankbaar voor hun medewerking. Dankzij jullie heb ik veel inzicht gekregen in het reilen en zeilen van een studie.

Daarnaast wil ik graag de volledige borstkliniek hartelijk bedanken. Prof. Dr. Tjalma, die de drijvende kracht was achter de vlotte organisatie van de studie, en die reeds op voorhand alle pijnpunten wist te tackelen. Nogmaals bedankt hiervoor! Prof. Dr. Peter Van Dam en Prof. Dr. Trinh, bedankt voor jullie kritische kijk op de zaken en voor jullie vlotte samenwerking in de staalverzameling. Dr. Christophe Van Berckelaer, bedankt om zo enthousiast mee te werken en patiënten te includeren. Verder ook bedankt aan alle ASO's (Leen, Osnat, Julie en Kelly) voor het opvolgen van de mails en de te includeren patiënten. Jullie hebben er mee voor gezorgd dat we zo weinig mogelijk inclusies misten. Ten laatste zou ik ook nog graag alle VTB's (Kelly, Nathalie, Ellen, Katleen en Yentl) willen bedanken om altijd zo behulpzaam te zijn en op al mijn vragen te antwoorden. Jullie zijn allemaal toppers!

Tijdens mijn project kwam ik bij een heel aantal diensten van het UZA terecht. Naar hen wil ik graag ook mijn dankbaarheid uiten, te beginnen met alle medewerkers van de biobank. In het bijzonder Pieter en Magali, bedankt om mij telkens verder te helpen. Ook alle laboranten, van harte bedankt voor jullie vele werk in het staaluitgifte- en verzamelingstraject. Verder wil ik graag het volledige MOCA-team bedanken, in het bijzonder Lise Verbruggen voor de vele efficiënte meetings en bij uitstek Sanne Wouters voor de leuke en vlotte samenwerking. Dankzij jullie is de COLIPAN een succes geworden, iets wat we bij de start niet hadden durven dromen. Bedankt aan jullie en alle andere medewerkers van het MOCA die me op weg geholpen hebben bij het opzetten van andere studies. Jullie advies en hulp is echt goud waard. Ten laatste ook bedankt aan het team van de Nucleaire Geneeskunde, in het bijzonder Dr. Ivan Huyghe, Dr. Laurens Carp en Dr. Taco Cappenberg om mij de kans te geven met de borstkankerpatiënten te spreken. Dit heeft gezorgd voor een vlotte workflow.

De laatste maar zeker niet de minste personen van het UZA die ik wil bedanken, zijn drie mensen die van onschatbare waarde zijn. Allereerst, Prof. Dr. Timon Vandamme. Aan de start van mijn doctoraat was ik slechts vaag op de hoogte van uw bestaan, maar dankzij de opstart van de maandelijkse meetings door Laura, kreeg ik het genoegen om u te leren kennen. Bedankt voor uw klinisch inzicht, voor uw hulp, al uw ideeën en uw

kritische vragen. Bedankt ook om mee uw schouders te zetten onder het COLIPAN-project en om zo enthousiast stalen te verzamelen. Daarnaast Dr. Dieter Peeters. Superdikke merci om alle stalen na te kijken, telkens met enthousiasme en kritische blik. Ik heb heel veel geleerd tijdens onze online en later ook live beoordelingssessies. Superhard bedankt om hiervoor tijd te willen maken. En last but not least, Els Vandersande. Beste Els, ik kan u niet genoeg bedanken voor alle hulp en moeite die u gedaan heeft voor mij. Dankzij u werd de bereikbaarheid van Prof. Peeters plots een stuk toegankelijker.

Tijd voor een switch van ziekenhuis! Naast goede samenwerkingen met het UZA, hebben we namelijk nog een prachtige samenwerking op het AZ Maria Middelares in Gent lopen. Deze is bij uitstek te danken aan Prof. Dr. Christof Vulsteke en Dr. Lieselot Croes. Aan jullie beiden, bedankt voor de vlotte samenwerking! Lieselot, ook heel hard bedankt voor alle extra moeite die je gedaan hebt, dankzij jou is deze staalverzameling een echt succes geworden.

Daarnaast zijn er een heel aantal mensen van de Belgische Virtuele Tumorbank, die gezorgd hebben voor een mooie uitbreiding van onze adenoom-weefselcollectie. Aan iedereen die hieraan meewerkte: bedankt, Thank you, Merci Beaucoup! Mede door de opstart hiervan creëerden we een samenwerking tussen ons team en dat van Prof. Dr. Sabine Tejpar van het UZ Leuven. Bedankt om te willen optreden als interne PI en mij zo het grootste aandeel stalen te kunnen bezorgen. Hierbij wil ik ook heel graag Bie bedanken, die mij doorheen het hele ethische traject van het UZ/ KU Leuven geloodst heeft, dit was geen makkelijke weg maar het is uiteindelijk gelukt!

Een groot deel van mijn doctoraat heb ik ook samengewerkt met de prachtige mensen van de SEHPO-onderzoeksgroep onder leiding van Prof. Dr. Guido Van Hal. Guido, ik zou u graag willen bedanken om vanaf de start van mijn doctoraat, zo vlot samen te werken. Ook al waren de beursaanvragen niet succesvol, we hebben er een goede connectie aan overgehouden die op andere tijdstippen van mijn doctoraat zeker van pas kwamen. Bedankt om aan mij te denken in samenwerkingen voor papers, bijvoorbeeld bij de review van Allegra of het pathologie project. U heeft mij veel kansen gegeven die tot mooie papers hebben geleid. Daarnaast wil ik ook graag Dr. Sarah Hoeck bedanken om altijd met zeer nuttige feedback klaar te staan op zo'n positieve manier. Bedankt voor de begeleiding! Furthermore, my gratitude goes to both Ngan and Allegra, who have crossed my path several times. Ngan, thank you first and foremost for the friendly interactions, for your great scientific knowledge and for your guidance in the pathology project. It was an honour to have worked with you so closely. Allegra, it was very nice to be able to work together with you! I am glad that after a long time of online meetings, we finally met on the Cancer Research Day. Your enthusiasm is catching. I hope to stay in touch with you both. Ten laatste zou ik ook graag Sharon en Koen van het Belgisch Kanker Register bedanken voor de fijne samenwerking in het laatste jaar van mijn doctoraat.

Zoals al duidelijk is gebleken, doe je een doctoraat niet alleen. Daarom wil ik graag ook even de aandacht vestigen op mijn fantastische collega's. Dear oncogenetic-colleagues, friends, I cannot explain how grateful I am for our group. I never dreamed of being part of such an amazing team. You are all golden and I care for all of you.

Liefste An, wat zouden wij toch zonder jou moeten beginnen. Naast al het harde werk ben jij voor ieder van ons een ongelofelijke steun en toeverlaat. Ik ben blij dat ik halverwege dit traject verhuisd ben en dat jij mijn bureau-buurvrouw bent geworden. Bedankt voor alle gezellige babbeltjes en om te luisteren en te helpen. Mijn doctoraat was nooit geworden wat het is zonder jou. Ik ben je hier heel erg dankbaar voor. Naast een fantastische grootmoeder voor Rhé, ben je ook een beetje een mama voor ons allemaal. Ik ga je missen. Gitta, jij bent reeds mijn inspiratie geworden toen ik met jou kwam praten op de onderzoeksdag van de faculteit bij de poster van het oncogenetica team. Ik heb altijd opgekeken naar jou, hoe je de meest kritische vragen kon stellen op zo'n vriendelijke en aanmoedigende manier. Bedankt hiervoor! Joe, you are such an impressive person and I am forever grateful I got to know you. I am proud to call you my colleague/friend. Thank you for always making time to check in with me and for all the analyses. My PhD could literally not have existed without you. Katleen, heel erg bedankt voor alle leuke momenten in het lab en daarbuiten. Onze samenwerkingen liepen altijd super, van het opstarten van een ddPCR assay tot arrays lopen. Ik ben heel blij dat we dit samen mochten doen. Bedankt om me zoveel bij te leren over de kliniek. Je bent een harde werker en dit werkt aanstekelijk voor ons allen. Joe and Katleen together, thank you for the nice teamwork, we were EPIC!! Our time in Madrid was the best time I had at a conference, I will cherish these moments forever. Lieve Janah, wat ben ik blij dat ik samen met jou het hele smMIPs traject heb meegemaakt. Onze samenwerking liep voor mij al van dag één vanzelf. Bedankt voor jouw nuchterheid en kalmte, hier kan ik veel van leren. Ik ben enorm dankbaar voor alle momenten in het labo. Thomas, merci voor al het werk en de leuke, 'spontane' koffiepauzes. Met jou erbij werd ons smMIPs-team compleet. Jullie beiden, we deden echt wel 'an IMPRESSive job'! Lisse, mercikes om vanaf het prille begin mij op te nemen in de groep. Ik ben blij dat ik naast jou kwam te zitten in de bureau! Merci voor alle leuke gesprekken toen en later, de leuke after-works en de gezellige niet-works. Laura, bedankt om de beste masterstudent ooit te zijn. Ik heb toen ook veel van jou geleerd. Het was superleuk om je nadien als collega te hebben. Bedankt voor al jouw harde werk en gezellige babbels. Danku dat ik altijd bij u terecht kan. Ana Regina, ik mag wel Ana zeggen hé, bedankt voor alle lieve attenties, toffe babbels en krullen-tips. Mijn congres-maatje, all-time goodie-scoorder, bedankt voor de fantastische tijden! Dear Marc and Farhan, you both started here during difficult times, but I am very grateful we got to know each other after the quarantine. Marc, thank you for your effortless kindness and for reminding us that there is more than just biomarkers in oncogenetic research. Farhan, thank you for broadening my cultural knowledge and for your humor. You will be a fantastic dad! Eef, bedankt voor al jouw inzet, de beste team-building ever en om ons de Nederlandse keuken te leren kennen (ik moet me eens een kaasschaaf aanschaffen). Nele, merci voor uw oneindig

enthousiasme, voor de wandelingetjes en de leuke uitstapjes. Vasilis, thank you for all your scientific insight and questions, I'm sure you are a great added value for our team. Arvid, bedankt om alles altijd te willen regelen, om mee van gedachten te wisselen als ik weer ergens vastzat en voor de leuke babbels. Bedankt om te willen luisteren als ik het ergens moeilijk mee had. Jij bent van onschatbare waarde op het CMG. Erik, bedankt om altijd tijd te maken voor mijn statistische vragen. Bedankt ook voor de leuke babbels en het enthousiasme op de afterworks.

Also, to all my other fantastic colleagues at CMG, thank you! Julia and Ligia, I am happy we could spend some time together and I appreciate all your valuable comments during Genetic seminars or brainstorm presentations, I have learned from you both. Furthermore, I would like to thank my former desk buddies Joe D, Eline, Dorien and Ilse L for the nice time we had. Verder mijn huidige bureau-buddies Laura R, Pauline, Maaike B, Sofie, Martina en Jarl. Bedankt voor alle toffe babbels en leuke afterworks. Maaike B en Jarl, bedankt dat ik altijd bij jullie terecht kon voor mijn lab probleempjes. Pauline en Laura R, merci voor de geweldig fijne babbels en after/outside work activiteiten! And all others of CMG: Yentl, Maaike A, Bert, Dale, Ilse VG, Melanie, Jolien H and S, Mathijs, Tammy, Irene, Lucia… thank you for the relaxed lunches, your kindness, the great afterworks and much more! Binnen het CMG zou ik graag nog NGS-team team bedanken om alle sequencing runs te verzorgen. Extra dankjewel hiervoor aan Ellen S, om de twee last-minute runs op te zetten. Bedankt ook aan het NIPT-team (Katrien, Els, Amber, Hilde en Ann) voor alle cfDNA extracties zo vlot te laten verlopen. Merci aan het team van het secretariaat om mijn staaltjes telkens te bezorgen en in het bijzonder veel dank aan Ann Van Bogaert voor alle hulp en de fijne babbels.

Naast het CMG zijn we als oncogenetica ook deel van het CORE en ik zou graag iedereen daar willen bedanken die op een of andere manier meegewerkt heeft aan dit project, in het bijzonder Delphine Quatannens. Bedankt om mij wegwijs te maken in de wereld van celcultuur en om de celletjes altijd te willen ontdooien en klaarzetten. Ze zijn een supergrote hulp geweest in al onze optimalisatie experimenten! Graag zou ik ook Evy Pluym willen bedanken voor alle hulp bij het opstarten en afronden van mijn doctoraat, en om tussendoor ook voor een goede administratieve begeleiding te zorgen.

Ook mijn vrienden zijn in deze periode van groot belang geweest. Eerst en vooral alle mensen van de middelbare school groep en de toevoegingen vanuit de KUL, jullie zijn er letterlijk en figuurlijk één uit de duizend! Lieve Marie en Jasmine, Nicholas en Daan, bedankt om er altijd te zijn wanneer het nodig was. Onze momenten zijn telkens lichtpuntjes in soms zeer stressvolle periodes. Lieve Charlotte/Lotje/Chisa, merci om er altijd te zijn en voor de overweldigende support in alles wat ik onderneem. Liefste Casper, bedankt voor alle leuke etentjes de afgelopen jaren en om met zoveel interesse naar mijn verhalen/ gezeur over het onderzoek te luisteren. Allerliefste Evelien, jou ben ik ontzettend dankbaar voor alle gesprekken, uitjes, belletjes, gewoon voor alles. Bedankt om altijd voor mij klaar te staan. Ten laatste gaat mijn eeuwige dankbaarheid uit naar Jotte.

Ik kan niet beschrijven hoe ongelofelijk fijn het was om na onze studies ook op 'hetzelfde' labo te belanden. Merci om mij mee uit mijn schulp te trekken, om mee te gaan wandelen als ik het even niet meer zag zitten, mij open te leren communiceren en om altijd te willen luisteren naar zelfs de idiootste stresskes, maar allermeest, bedankt om jezelf te zijn.

Bedankt aan mijn lieve familie en schoonfamilie om mij aan te moedigen en interesse te tonen in iets wat toch wel een ver-van-jullie-bed-show is. Merci Bon-Papa, Oncle Jean, Parrain, Tante Myriam, Elise, Tonton Vincent, Tantine, Mélanie et Ignace pour vos mots d'encouragement. Bedankt lieve marraine, oncle Pierre, Julie en Camille voor jullie bemoedigende woorden. Bedankt Elise & Jelle, Charlotte & Olivier, Mats & Katarina, en vooral Hilde en Kurt om altijd te willen luisteren naar mijn (waarschijnlijk onverstaanbare) verhalen over het werk. Bedankt liefste Mona, Stella en Otto om de allerleukste afleiding te zijn. Ook bedank ik graag zij die er niet meer zijn. Merci ma chère Bonne-Maman, d'être fière de moi. Dankjewel mijn liefste omaatje, je was mijn inspiratie om aan dit onderzoek te starten en ik heb erg veel aan je gedacht. Ook bedankt aan mijn lieve opa. Bedankt om in mij te geloven en duidelijk erg trots te zijn op alles wat ik bereikt heb. Ik mis jullie.

Uit de grond van mijn hart bedank ik mijn allerliefste gezin. Christophe'ke', mijn kleine grote broer, bedankt om altijd te luisteren naar mijn stresskes, en heel erg bedankt voor het prachtige design van mijn boekje! Sophie, bedankt om er voor mij te zijn al vanaf mijn studies. Dank u dat ik altijd mocht binnenvallen met mijn paniekskes en dank u om ook tijdens mijn doctoraat naar al mijn zorgen en verhalen te luisteren. Danke schön ook Fabian, om zo begaan te zijn met mij en mijn werk. Liefste mama en papa, ik kan amper beschrijven hoe dankbaar ik jullie ben. Dankzij jullie is dit allemaal mogelijk geweest. Bedankt voor jullie overweldigende steun, jullie motiverende woorden, om te luisteren naar al mijn (stomme) zorgen en om mij te kalmeren wanneer ik het eens niet zag zitten. Ik had hier vandaag niet gestaan zonder jullie.

Graag doe ik ook een kleine shout-out naar mijn allerliefste Toulouse (Catito), die ondanks dat ze het zelf totaal niet doorheeft, mij zo hard helpt met al haar lieve knuffeltjes. En laatst maar zeker niet allerminst gaat al mijn dankbaarheid uit naar mijn toekomstige man, Nils. Mijn lieve schat, ik zou dit nooit gekund hebben zonder jouw eindeloze geduld, je steun en je kalmerende woorden. Dankjewel om te luisteren naar mijn frustraties en om de dingen te relativeren als ik door de bomen het bos niet meer zag. Dankjewel dat ik jou altijd kon inschakelen als klankbord over mijn onzekerheden. Je bent er heel mijn doctoraat lang voor me geweest alsof het vanzelfsprekend was (ook al is het dat zeker niet altijd). Merci om mij op de juiste momenten enthousiast te maken/houden, en om mij af te remmen als ik overdrijf. Jij bent mijn alles, ik hou van jou.

Oneindig bedankt iedereen, voor alles!

Isabelle

Promoters: Prof. Dr. Ken Op de Beeck | Prof. Dr. Marc Peeters | Prof. Dr. Guy Van Camp

Mentor: Prof. Dr. Wiebren Tjalma

Proefschrift voorgelegd tot het behalen van de graad van Doctor in de Biomedische Wetenschappen

Faculteit Farmaceutische, Biomedische en Diergeneeskundige Wetenschappen - Departement Biomedische Wetenschappen

Antwerpen, 2024

University of Antwerp