U.PORTO

MESTRADO ONCOLOGIA

ESPECIALIDADE EM ONCOLOGIA LABORATORIAL

Cell-free DNA methylation as a biomarker for early detecting the four major cancers

Tiago Brito da Rocha

M

2022

U.PORTO

Tiago Brito da Rocha. Cell-free DNA methylation as a biomarker for early detecting the four major cancers

M.ICBAS.2022

Cell-free DNA methylation as a biomarker for early detecting the four major cancers

Tiago Brito da Rocha

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

Tiago Brito da Rocha

Cell-free DNA methylation as a biomarker for early detecting the four major cancers

Dissertação de Candidatura ao grau de Mestre em Oncologia – Especialização em Oncologia Laboratorial submetida ao Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto

Orientadora: **Professora Doutora Carmen de Lurdes Fonseca Jerónimo**
Professora Catedrática Convidada
Departamento de Patologia e Imunologia Molecular
Instituto de Ciências Biomédicas Abel Salazar - Universidade do Porto
Investigadora Auxiliar e Coordenadora do Grupo de Epigenética e Biologia do Cancro
Diretora do Centro de Investigação
Instituto Português de Oncologia do Porto Francisco Gentil, E.P.E

Coorientador: **Professor Doutor Rui Manuel Ferreira Henrique**
Professor Catedrático Convidado
Departamento de Patologia e Imunologia Molecular
Instituto de Ciências Biomédicas Abel Salazar - Universidade do Porto
Assistente Graduado Sénior do Serviço de Anatomia Patológica
Investigador Sénior do Grupo de Epigenética e Biologia do Cancro
Centro de Investigação
Instituto Português de Oncologia do Porto Francisco Gentil, E.P.E

This study was funded by the Research Center of Portuguese Oncology Institute of Porto (Grants CI-IPOP-74-2016-MethylBiom4Can and CI-IPOP-27-2016-FBCBEG) and approved by the institutional review board (Comissão de Ética para a Saúde – CES-IPOFG 120/015)

Agradecimentos

Esta dissertação não seria possível sem o contributo de várias pessoas às quais não podia deixar de agradecer. Foram 2 anos de muito crescimento não só científico, mas também pessoal, graças a todos vocês!

Em primeiro lugar, um grande obrigado à minha orientadora, Professora Doutora Carmen Jerónimo, por me ter permitido desenvolver este projeto no Grupo de Epigenética e Biologia do Cancro (GEBC). Obrigado pelo voto de confiança e por todo o conhecimento partilhado, bem como toda a disponibilidade, mesmo quando o pouco tempo livre não estava a favor! E ao meu coorientador, Professor Doutor Rui Henrique, por toda a sua contribuição imprescindível neste trabalho.

A todos os pacientes, dadores de sangue e profissionais de saúde envolvidos na colheita das amostras utilizadas, um sincero obrigado! Este trabalho não seria possível sem a vossa gentil contribuição!

À Vera, todos os obrigados do mundo não seriam suficientes para expressar o quão grato estou por toda a tua ajuda! Obrigado por tudo o que me ensinaste, desde trabalho laboratorial à capacidade crítica e, principalmente, pela tua disponibilidade, pois mesmo estando longe nunca me negaste uma chamada ou uma reunião de última hora, fosse por que razão! Este trabalho é tanto teu como meu! Muito obrigado!

À Margareta, pelo teu sorriso contagiante e por essa energia vibrante! Obrigado por partilhares toda a tua sabedoria e experiência de vida e mostrares que fazer da ciência vida não é assim tão mau quanto parece... À Bianca, minha parceira ressolha, pelo teu sorriso matinal diário, pelas horas astrológicas e por acordares o pt br dentro de mim! À Pipita e a sua energia pacífico-caótica e interjeições sempre muito oportunas! Todos os almoços, cafés, lanches, jantares (dias inteiros no geral) ao vosso lado fizeram de mim uma pessoa muito feliz nos últimos tempos! Este ano foi especial graças a vocês e sei que tenho 3 lindas pessoas com quem poderei sempre contar. Obrigado pela vossa essência!

À Bia, por me acompanhar nesta jornada desde o início. Olhar para trás e ver o quanto crescemos juntos deixa-me um sorriso gigante na cara. Já mais vou esquecer todas as horas passadas na cultura com as devidas playlists, todos os desabafos, todas as frustrações, todos os bons momentos, toda a partilha, tudo! Obrigado pela pessoa que és e pela pessoa que me fazes ser. Ansioso por te ver brilhar no futuro e por continuar a crescer ao teu lado!

Aos restantes membros do GEBC, por não hesitarem em esclarecer qualquer dúvida que me surgisse e por me terem acolhido no seu ambiente de trabalho. Muito obrigado a todos pela vossa ajuda!

À Carolina, Tânia, Diogo e Miguel, obrigado por terem tornado este mestrado mais especial e pelos jantares extremamente necessários para manter a nossa sanidade mental, embora não ótimos para a nossa saúde...

À Ana, Ana Costa, Gui, Hugo, Mafalda e Teresa, obrigado por serem a minha segunda família. Os últimos 5 anos ao vosso lado foram sem dúvida os melhores anos da minha vida e não podia pedir melhor que vocês. Obrigado pela vossa essência, pela vossa individualidade e pela vossa amizade. Crescer com vocês foi um prazer e será ainda melhor continuar a fazê-lo!

À Sabrina, minha alma gémea e terapeuta pessoal. Obrigado por seres o meu escape da vida séria e pelas nossas aventuras! Obrigado pela pessoa que és, por todos os incentivos e por acreditares sempre em mim!

Por fim, mas não menos importante, à minha família por sempre confiarem em mim e me incentivarem a seguir os meus sonhos! Obrigado por me permitirem estar onde estou e ser quem sou.

Muito obrigado a todos!

Resumo

Introdução: Os cancros da mama (CaMa), pulmão (CaPl), coloretal (CaCr) e próstata (CaP) são os quatro cancros mais incidentes a nível mundial para ambos os sexos. Apesar de os métodos de rastreio disponíveis terem contribuído para reduzir a incidência e a mortalidade, existe ainda uma necessidade de desenvolver novas formas eficazes e minimamente invasivas de detetar precocemente estas neoplasias. A hipermetilação do promotor de genes é um evento precoce na carcinogénese, sendo específica quer de tumores, quer do tipo de tecido, e facilmente obtida de fluídos corporais, sendo então um potencial biomarcador para uma deteção precoce e minimamente invasiva destes cancros. Uma análise de dados da TCGA foi realizada para encontrar promotores de genes hipermetilados especificamente em cancro e foram identificados os promotores do *ADCY4, MIR129-2, NID2* e *MAGI2* comumente hipermetilados no CaMa, CaPl, CaCr e CaP. Assim, o objetivo deste trabalho foi desenvolver um teste baseado na metilação do DNA livre em circulação (cfDNA) para deteção simultânea destes quatro cancros.

Métodos: Ensaios de *droplet digital PCR* (ddPCR) específicos para metilação foram desenvolvidos de modo a avaliar os níveis de metilação do promotor dos genes mencionados acima. cfDNA foi extraído de amostras de plasma de 50 pacientes com CaMa, 50 com CaPl, 50 com CaCr e 50 com CaP e 50 doadores de sangue, seguido por modificação bissulfito. Os níveis de metilação do promotor do *ADCY4, MIR129-2, NID2* e *MAGI2* foram avaliados através dos ensaios de ddPCR desenvolvidos.

Resultados: Todos os promotores apresentaram níveis de metilação significativamente elevados em pacientes com cancro relativamente aos controlos, exceto o *MIR129-2* e *NID2* em pacientes com CaMa. Um painel "PanCancer" combinando todos os genes e, considerando uma amostra positiva sempre que um gene é positivo, detetou CaMa com 52% de sensibilidade, CaPl com 85.71%, CaCr com 78.72% e CaP com 56.41% a uma especificidade de 93.75%. A deteção do cancro em estadios iniciais mostrou resultados similares a todos os estadios. O painel "PanCancer" apresentou 69.80% de sensibilidade, 93.75% de especificidade e 74.03% de precisão na deteção simultânea dos quatro cancros.

Conclusões: O teste "PanCancer" baseado na metilação do DNA em biópsias líquidas apresentou um grande potencial para a deteção precoce dos CaMa, CaPl, CaCr e CaP, oferecendo uma nova ferramenta de pré-triagem de doentes para o exame diagnóstico mais adequado, permitindo aumentar a adesão aos rastreios e reduzir os custos nos sistemas de saúde.

Abstract

Background: Breast (BrC), lung (LC), colorectal (CRC) and prostate (PCa) cancers are the four most incident cancers worldwide for both males and females. Despite available screening methods having contributed to incidence and mortality reduction, there is still a great need for the development of minimally invasive and effective tools for early detecting these malignancies. Gene promoter hypermethylation is an early tumorigenic event, being cancer- and tissue-specific and easily obtained from body fluids, thus being a potential biomarker for minimally invasive pan-cancer early detection. A TCGA data mining was performed to select cancer-specific hypermethylated gene promoters depicting *ADCY4, MIR129-2, NID2* and *MAGI2* promoters as commonly hypermethylated in BrC, LC, CRC and PCa. Herein, we aimed to develop a cell-free DNA (cfDNA) methylation-based test, using the above targets, for simultaneously detecting the four major cancers.

Methods: Methylation-specific droplet digital PCR (ddPCR) assays were developed for assessing gene promoter methylation levels. cfDNA was extracted from plasma samples of 50 BrC, 50 LC, 50 CRC and 50 PCa patients and 50 healthy blood donors, followed by sodium-bisulfite modification. *ADCY4, MIR129-2, NID2* and *MAGI2* promoter methylation levels were evaluated using the developed ddPCR assays.

Results: All gene promoters displayed significantly higher methylation levels in cancer patients compared to controls, except *MIR129-2* and *NID2* in BrC patients. A “PanCancer” panel combining all genes was built and considering a positive sample whenever one gene was positive, we detected BrC with 52% sensitivity, LC with 85.71%, CRC with 78.72% and PCa with 56.41% at 93.75% specificity. Early-stage cancer detection showed similar performance to all-stage cancer. The “PanCancer” panel depicted 69.80% sensitivity, 93.75% specificity and 74.03% accuracy for simultaneous detection of the four cancers.

Conclusions: The “PanCancer”, a DNA methylation-based test in liquid biopsies showed great potential for early detecting BrC, LC, CRC and PCa, unveiling a new tool for pre-screening patients for further adequate diagnostic examinations, hopefully, increasing compliance to screening and reducing the costs on health care systems.

Table of Contents

Figures Index

Tables Index

List of Abbreviations

5hmC – 5-hydroxymethylcytosine

5mC – 5-methylcytosine

A

AC – Asymptomatic control

ADC – Adenocarcinoma

ADCY4 – Adenylyl cyclase type 4

ADCY4me – Methylated adenylyl cyclase type 4

ADT – Androgen deprivation therapy

ALK – ALK receptor tyrosine kinase

AOX1 – Aldehyde oxidase 1

APC – APC regulator of WNT signalling pathway

AR – Androgen receptor

AUC – Area under curve

B

BRAF – B-Raf proto-oncogene, serine/threonine kinase

BrC – Breast Cancer

C

CDO1 – Cysteine dioxygenase type 1

CE-IVD – *Conformité Européenne* – In Vitro Diagnostic

CELF2 – CUGBP Elav-like family member 2

cfDNA – cell-free DNA

cfRNA – cell-free RNA

CHFR – Checkpoint with forkhead and ring finger domains

CIMP – CpG island methylator phenotype

CIN – Chromosomal instability

CP – Normal prostate tissue

CpG – Cytosine-phosphate-guanine

CRC – Colorectal Cancer

CRN – Normal colorectal tissue

CT – Computed tomography

CTC – Circulating tumor cell

ctDNA – Circulating tumor DNA

CXR – Chest X-Ray

D

DBT – Digital breast tomosynthesis

DCIS – Ductal carcinoma in situ

ddPCR – Droplet digital polymerase chain reaction

DNMT – DNA methyltransferase

dPCR – Digital polymerase chain reaction

DRE – Digital rectal examination

E

EDNRB – Endothelin receptor type B

EDTA – Ethylenediaminetetraacetic acid

EGFR – Epidermal growth factor receptor

ER – Estrogen receptor

ETS – E26 transformation-specific

EV – Extracellular vesicle

F

FAP – Familial adenomatous polyposis

FDA – US Food and Drug Administration

FFPE – Formalin-fixed paraffin-embedded

FIT – Fecal immunochemical test

FLOT1 – Flotillin 1

FOBT – Fecal occult blood test

G

GFRA1 – GDNF family receptor alpha-1

H

HDAC – Histone deacetylase

HER2 – Human epidermal growth factor receptor 2

HNPCC – Hereditary non-polyposis colon cancer

HOXA11 – Homeobox A11

I

IDC – Invasive ductal carcinoma

IHC – Immunohistochemistry

ILC – Invasive lobular carcinoma

K

Ki-67 – Marker of proliferation Ki-67

KRAS – Kirsten rat sarcoma viral oncogene homolog

L

LC – Lung cancer

LCIS – Lobular carcinoma in situ

LCLC – Large cell lung cancer

LDCT – Low-dose computed tomography

LOB – Limit of blank

LOD – Limit of detection

LOQ – Limit of quantification

M

MAGI2 – Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2

MAGI2me – Methylated membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2

MBP – Methyl-binding protein

MCED – Multi-cancer early detection

MIR129-2 – microRNA-129-2

MIR129-2me – Methylated microRNA-129-2

MRI – Magnetic resonance imaging

MSI – Microsatellite instability

N

NBr – Normal breast tissue

NID2 – Nidogen 2

NID2me – Methylated Nidogen 2

NL – Normal lung tissue

NSCLC – Non-small cell lung cancer

P

p40 – Isoform of p63 (deltaNp63)

p53 – Tumor suppressor p53

PCa – Prostate cancer

PCR – Polymerase chain reaction

PIK3CA – Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha

PR – Progesterone receptor

PSA – Prostate-specific antigen

PTGER4 – Prostaglandin E receptor 4

Q

qMSP – Quantitative methylation-specific polymerase chain reaction

R

ROC – Receiver operating characteristic

ROS1 – Proto-oncogene tyrosine-protein kinase ROS

S

SCC – Squamous cell carcinoma

SCLC – Small cell lung cancer

SEPT9 – Septin 9

SHOX2 – Short-stature homeobox 2

SSP – Sessile serrated polyp

T

TEP – Tumor-educated platelet

TET – Ten-eleven translocation methylcytosine dioxygenases

TOO – Tissue-of-origin

TTF-1 – Thyroid transcription factor 1

U

USPSTF – United States Preventive Services Taskforce

W

WHO – World Health Organization

Z

ZSCAN1 – Zinc Finger And SCAN Domain Containing 1

I.INTRODUCTION

1. Four major cancers

According to GLOBOCAN 2020, considering both genders, breast cancer (BrC), lung cancer (LC), colorectal cancer (CRC) and prostate cancer (PCa) are the most incident cancers worldwide. In fact, these four cancers account for over 40% of all cancer diagnoses and about 38% of all cancer deaths, worldwide (Figure 1) [1].

In males, LC is the most diagnosed cancer, followed by PCa and CRC. Moreover, LC is also the leading cause of death by cancer in males, followed by liver cancer and CRC. Among females, BrC is the most diagnosed malignancy, followed by CRC and LC in incidence and mortality [1].

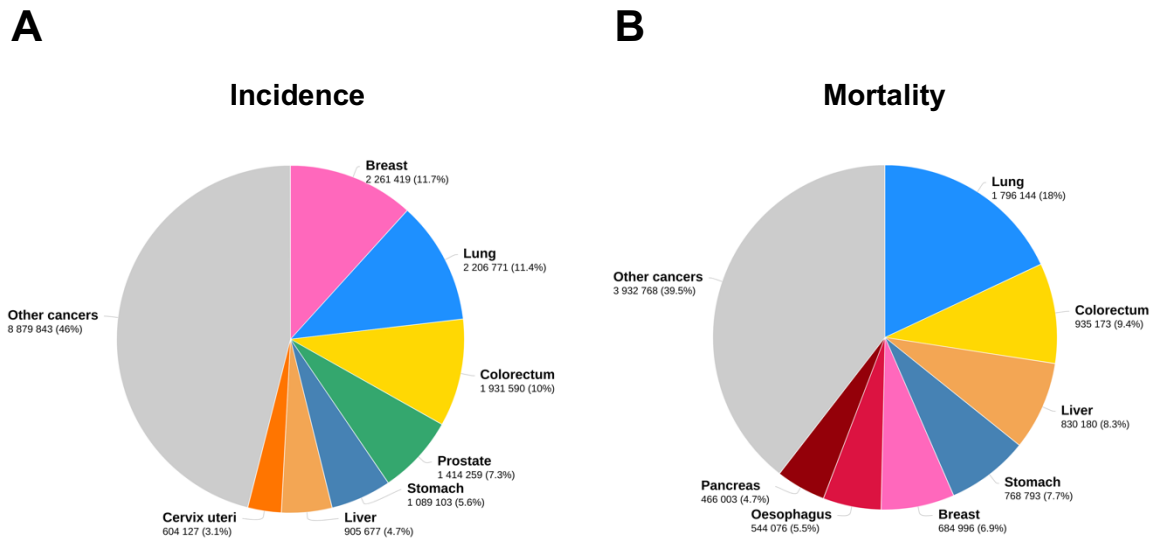

Figure 1. Pie charts representing the percentage of cancer-related incidence (A) and mortality (B) in both sexes, worldwide, in 2020. Adapted from [1].

1.1 Breast Cancer

Worldwide, 2.3 million new BrC cases were estimated in 2020, which accounts for 1 in 4 cancer cases in women [1]. BrC can also occur in men, but it is more than 100 times more common in females than in males [2]. Usually, developed countries present higher BrC rates compared to developing countries, which may be due to certain lifestyle and reproductive factors more common in developed countries. However, with the ‘‘Westernization’’ of developing countries, BrC incidence rates have been rising over the last decades, leading to a global incidence increase [1, 3]. BrC is most common in women aged 55 to 64 and is rarely diagnosed in women younger than 40. As expected, the risk increases with age, although decreasing after menopause, suggesting that hormones play a crucial role in breast carcinogenesis [3]. In fact, environmental, lifestyle and reproductive factors have

been shown to have impact in BrC: late menarche and early menopause are associated with less risk of BrC and it has also been shown that postmenopausal BrC risk reduces by 4% for each year that menarche is postponed [3]; women who went through pregnancy have reduced risk of BrC and studies revealed that each pregnancy reduces premenopausal BrC risk by 3% and by 12% for postmenopausal BrC. Also, earlier first births and increased number of births have been associated with lower risk [3], along with breastfeeding and increased breastfeeding [3, 4]; Contrarily, oral contraceptive use seems to be linked with increased BrC risk, however, research in this topic is not enough to fully support this association [3, 4]. Additionally, family history of breast cancer, especially in one or more first-degree relatives is a strong risk factor for BrC [4]. Although with less consistency, external factors, such as alcohol consumption, diet, especially with high fat intake, obesity and exposure to radiation are also considered risk factors for BrC development. On the other hand, physical activity has been associated with BrC risk reduction [3].

BrC has been recognized as a heterogeneous disease for a long time, with a large degree of inter- and intra-tumor variability [4]. It can occur in any cell of the mammary gland and displays a variety of morphologies, IHC profiles and histological subtypes, each leading to specific clinical outcomes [2]. Carcinomas account for the majority of BrC and are histologically classified according to two factors: if the tumor is limited to breast epithelium (in situ carcinoma) or invaded the stroma (invasive carcinoma); if the tumor originates from the ducts (ductal carcinoma) or from the lobules (lobular carcinoma). Hence, BrC can be subdivided in ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC), lobular carcinoma in situ (LCIS) and invasive lobular carcinoma (ILC) [2]. IDC is the most common form of BrC with 55% incidence, followed by ILC with 5%-15% incidence. [2]. Importantly, molecular classification provides information for targeted therapies and allows for individualized treatment [5]. According to different genes' expression, BrC is classified in luminal A, luminal B, HER2 and basal-like. Luminal A is positive for estrogen (ER) and progesterone receptors (PR) expression and negative for human epidermal growth factor receptor-type 2 (HER-2) expression. Luminal B is also positive for ER/PR, meanwhile, expression of HER-2 may vary. It is distinguished from luminal A due to increased Ki-67 expression. HER2 is negative for ER/PR and is characterized by a highly positive HER-2 expression. Basal-like or tiple-negative is negative for ER/PR and HER-2 [2, 5].

1.1.1 Screening and diagnosis

Early BrC diagnosis without detectable metastasis is a potentially curable disease [6]. Therefore, effective screening and early detection are essential for mortality reduction.

Mammography is the gold-standard screening method for BrC and it significantly reduces the risk of death from the disease. Data from randomized control trials have shown that screening mammography reduces BrC cancer mortality by at least 20% [7]. Nowadays, several European countries have national or regional mammography screening programs aiming to detect BrC at an early stage and recommend this procedure to be taken every two years. In fact, a two-year time interval between screening allowed for mortality reduction in women between 50 and 69 years old [8]. Importantly, annual screening in women aged between 40 and 84 years has shown more impressive results in mortality reduction [7]. Indeed, the guidelines for BrC screening are not consensual and different organizations suggest different time and age intervals [7, 9]. Digital breast tomosynthesis (DBT), ultrasound, and magnetic resonance imaging (MRI) are other methods for BrC screening that appear to improve its detection [7]. For instance, DBT enables BrC detection even in women with dense breasts, a factor that is known to interfere with mammography accuracy [10]. However, these procedures can be more expensive and require more samples, being more time consuming [7, 10]. Besides its benefits in mortality reduction, BrC screening also presents some disadvantages such as overdiagnosis [11]. False positive results, unnecessary biopsies, and radiation exposure are also harmful consequences of BrC screening [9, 11].

Currently, BrC diagnosis is based on clinical examination, such as breasts and lymph nodes palpation, and is confirmed by histopathological evaluation of tissue samples obtained by core needle biopsy from the primary tumor or metastasis, if suspected [8]. Personal medical history, breast/ovarian cancer-related family history, and menopausal status assessment are other factors usually considered for diagnosis [8].

1.2 Lung Cancer

Worldwide, 2.2 million new LC cases and 1.8 million fatalities were estimated in 2020 [1]. The World Health Organization (WHO) estimates that LC is the cause of 1.59 million deaths globally per year [12], which happens mainly because it is mostly asymptomatic, leading to a late-stage diagnosis [13].

Tobacco smoking is an important LC risk factor, being responsible for 80-90% of all LC cases. However, only 15% of smokers develop LC, which suggests the relevance of other factors, namely genetic/epigenetic factors. Males appear to be more susceptible to develop LC than females, although, large cohort studies have found no association between gender and risk of LC. In addition, LC usually does not manifest until people reach around 50 years old and the risk of developing the disease increases with age, thereafter, being the median age of diagnosis at 70 years [13, 14].

Based on histology, LCs are classified in two main groups: small-cell carcinoma (SCLC), which accounts for 10-15% of all cases, and non-small-cell carcinoma (NSCLC), the most common form of LC, accounting for 85-90% of all cases [15]. NSCLC encompasses adenocarcinoma (ADC), squamous cell carcinoma (SCC) and large-cell lung carcinoma (LCLC). ADC accounts for about 40% of all LC cases, being, in fact, the major diagnosed subtype, followed by SCC with 20% of incidence and, lastly, LCLC, which represents a small percentage of about 3% of all LCs [15]. LCLC is also known as undifferentiated NSCLC, due to the fact that its diagnostic is achieved by ruling out the other three subtypes [15]. SCLC is the most tobacco smoking associated LC type, with 95% of all diagnosed cases having a history of tobacco exposure. Since smoking rates tend to be declining, this might explain why SCLC incidence is lower compared to NSCLCs [16].
In addition to histological classification, genetic alterations, such as *EGFR* mutations and *ALK* and *ROS1* rearrangements provide prognostic information allowing for patient tailored therapy [15].

1.2.1 Screening and diagnosis

As mentioned, LC has a high mortality rate since it is usually detected at late stages, when patients present symptoms. Therefore, screening is crucial for early diagnosis and mortality reduction [13, 17].

Currently, low-dose computed tomography (LDCT) and chest X-ray (CXR) are two available screening methods. In the USA, in a study conducted by the National Lung Screening Trial, researchers demonstrated a 20% reduction in LC mortality by LDCT compared with CXR, which only decreased mortality by 6.7%. However, 96.4% of the positive results by LDCT and 94.5% of those by CRX were false positive results [18]. Therefore, the high false positive rates raised by these screening methods are a major concern, in addition to the excessive radiation exposure [17]. Nonetheless, the US Preventive Services Task Force (USPSTF) recommends annual screening by LDCT in individuals aged 50–80 at increased risk of developing LC i.e., individuals with >20 pack-year smoking history, currently smoking or that have quit within the last 15 years [19].

Regarding LC diagnosis, the current gold-standard procedure is microscopic evaluation of histological or cytological samples, providing the confirmation of the tumor's presence and its histological type, based on morphology and IHC. Samples are usually obtained from the primary tumor or metastasis through a bronchoscopy [12, 15]. IHC serves primary diagnosis, as well as prediction of LC subtypes according to specific markers: positive expression of thyroid transcription factor 1 (TTF-1) favors ADC diagnosis, positive expression of p40 favors SCC diagnosis and both markers are usually used in combination.

If the expression is negative for both, the diagnosis remains NSCLC-not otherwise specified (NSCLC-NOS), due to its inconclusiveness [12].

1.3 Colorectal Cancer

Over 1.9 million new CRC cases and 935,000 deaths were estimated in 2020, for both genders, accounting for 1 in 10 cancer cases and related deaths [1]. Incidence is low at ages younger than 50 years, but increases with age, being the median age of diagnosis around 70 years in developed countries [20]. Also, the risk seems to be about 2 times higher in men than in women [21].

Several environmental and lifestyle risk factors have been associated with CRC development, including high consumption of red and processed meats, high consumption of alcohol, increased body fat and diabetes [20]. Other pathologies, like ulcerative colitis and Crohn's disease were also associated with CRC development [21]. Evidence suggests that infection with *Helicobacter pylori*, *Fusobacterium spp* and other infectious agents might also be associated with CRC increased risk [20]. Contrarily, physical activity has been shown to be a protective factor [20]. Importantly, CRC also has a heavy heritable component. In fact, 35% of CRC risk might be due to heritable factors [20]. Over the last decades, genetic bases of several CRC syndromes have been identified, such as the Lynch syndrome or hereditary nonpolyposis colorectal cancer (HNPCC) and familial adenomatous polyposis (FAP), the two main heritable syndromes. Lynch syndrome is caused by mutations in genes involved in DNA mismatch repair and FAP is caused by *APC* gene mutations. In fact, *APC* gene mutations are an early event in CRC formation and occur in more than 70% of cases [20, 21]. Nonetheless, most CRC cases are sporadic i.e., are developed due to external factors that generate a sequential accumulation of somatic mutations that lead to genomic instability by 3 different pathways: chromosomal instability (CIN) pathway, caused by mutations in genes that activate the WNT pathway; microsatellite instability (MSI) pathway, due to loss or failure of DNA mismatch repair genes; CpG island methylator phenotype (CIMP) pathway, characterized by promoter hypermethylation and consequent silencing of many key tumor suppressor genes [21, 22].

Heritable or sporadic, CRCs develop slowly via a multistep process involving a series of histological, morphological, and genetic/epigenetic changes that accumulate over time. It typically develops from benign abnormal tissue growths, also known as polyps, within the intestinal mucosa, from which the cells start to proliferate and accumulate genetic and epigenetic changes, acquiring the ability to invade the lumen and eventually spread to distant metastatic sites [23]. There are two types of polyps with malignant potential, adenomas and sessile serrated polyps (SSPs), and each use different pathways to develop

into a tumor. For instance, adenomas commonly follow the so-called “adenoma–carcinoma sequence”, which consists of sequential mutations in the *APC* gene, followed by the *KRAS* oncogene and the *p53* gene. However, development of SSPs often begins with mutations in the *BRAF* gene and may also occur in the *KRAS* oncogene but are less frequent [20, 23].

1.3.1 Screening and diagnosis

While early stage and localized CRC display a 5-year survival rate in 90% of cases, those diagnosed at late stage and that have metastasized to distant sites only show a survival rate of about 13%. At late stages, CRC treatment is palliative and highly expensive, hence, effective screening methods are essential for mortality and cost reductions [23].

Colonoscopy is the current reference method for CRC screening and is recommended every 10 years in patients aged 50 years or older, presenting up to 95% sensitivity and nearly 100% specificity [23, 24]. It allows the visualization of the entire colon and the detection of both cancerous and precancerous lesions [23]. One of colonoscopy’s greatest advantages is that it allows the removal of polyps at the time of detection [23, 24]. However, some disadvantages include the invasiveness of the procedure and the risk associated with anesthesia and the required bowel preparation. Bowel preparation is often time-consuming due to the necessary temporary change in medications and diet and entails use of a cleansing agent, leading to electrolyte abnormalities. Perforation of the bowel and post-colonoscopy bleeding is also possible, but very rare [23, 24]. Sigmoidoscopy is another screening method available, very similar to colonoscopy, but less common, that only analyses the distal half of the colon. Besides, it does not require sedation and the bowel preparation is simply an enema on the day of the examination. However, if lesions are identified in the distal colon, a follow-up colonoscopy is required [23]. Computed tomography (CT) colonography consists in a structural radiologic examination of the colon that takes multiple pictures and allows to create a 2- or 3-dimensional image of the colon [23, 24]. Usually, this procedure is undertaken by patients who are not fitted for colonoscopy and has the disadvantage of radiation exposure [23]. Fecal occult blood testing (FOBT) and fecal immunochemical testing (FIT) are also screening methods available that detect hemoglobin as a marker of blood in feces [23]. These tests are non-invasive and do not require bowel preparation, although usually lead to false-positive results and have limited sensitivity to detect precancerous lesions [23]. While FOBT displays 79.4% sensitivity and 86.7% specificity, FIT can detect CRC with 79% sensitivity and 94% specificity, although only showing 28% sensitivity for detecting advanced adenomas [25].

For diagnosis, tissue samples obtained from the colonoscopies are further analyzed histologically by a pathologist [20, 26].

1.4 Prostate Cancer

Being the second most incident cancer among men, almost 1.4 million new PCa cases and 374,000 deaths were estimated in 2020 [1]. PCa is more commonly diagnosed in older men, being the average age at diagnosis 65 years. The risk for developing the disease increases after the age of 50 and it is expected that 1 in 6 men will be diagnosed with PCa during lifetime [27]. African American, African Caribbean and South American men also have higher incidence and mortality rates of PCa than Caucasian men [27]. Aside from age and race, the most well-established risk factor for PCa is family history of the disease. First-degree relatives of men with PCa have two times more risk for PCa than the general population and four times more risk if they are first-degree relatives of men diagnosed with PCa at age younger than 60 years [28]. Lifestyle factors, such as diet, smoking and alcohol consumption also seem to be associated with an increased risk for PCa [27].

The majority of prostate tumors are carcinomas, although, rarely, they can also be diagnosed as sarcomas or lymphomas [29]. Prostate carcinomas can be further subdivided in intraductal carcinomas, adenocarcinomas, adenosquamous carcinomas, squamous cell carcinomas and basal cell carcinomas, being adenocarcinomas the most commonly found tumors [29]. Adenocarcinomas might originate from prostatic intraepithelial neoplasms, that develop into localized adenocarcinomas and, eventually, into advanced adenocarcinomas and metastasis [30]. PCa can also be classified molecularly, according to specific genes expression, as ETS-positive or ETS-negative, based on the fusion of ETS genes with androgen-regulated, prostate-specific genes, the most common rearrangement that occurs in prostate carcinogenesis [31]. These rearrangements seem to occur early and might result from activated androgen receptors (AR) generating DNA damage through transcription at androgen-receptor binding sites [28]. AR are transcription factors that bind to testosterone and other androgens, becoming active, and then bind to the promoter region of their targeted genes to promote transcription. They are required for prostate normal functioning, however, their activation is also associated with PCa development, since several molecular alterations that occur in early stages of PCa involve genes that are androgen-dependent [31]. Hence, androgen deprivation therapy (ADT) is a common procedure for PCa treatment, and its application has led to a significant reduction in mortality, even in metastatic PCa [28, 31].

1.4.1 Screening and diagnosis

PCa screening is essential for identifying men with potentially lethal tumors during an asymptomatic phase, hopefully, leading to an increase in survival rates. Currently, the most widely used screening procedures are digital rectal examination (DRE) and serum prostate-

specific antigen (PSA) testing [32]. During many years, DRE was the primary screening test for PCa, however, the majority of cancers were detected, already, at an advanced stage [33]. In 1986, the PSA test was approved by the FDA as a tool for PCa patients monitoring but was rapidly converted in a screening test and widely used due to its ability of increased early detection of PCa, compared to DRE [33, 34]. PSA is a protease produced by prostatic epithelial cells and its levels are elevated in PCa, but also in other pathologies, such as prostatic hyperplasia or infections, which may lead to false-positive results [34]. Overdiagnosis of PCa is another drawback of PSA screening, with a study having reported 29% overdiagnosis for white men and 44% for black men [35]. Moreover, a normal PSA value does not exclude the existence of PCa. It was shown, in a clinical trial, that PCa was detected in 15% of men with normal results on PSA and also DRE who underwent a prostate biopsy at the end of the study [33]. Concerning screening guidelines, most guidelines do not recommend screening in men older than age 75, since the risks of adverse effects from treatment, comorbidity and overdiagnosis are increased. Intervals between screening should be 2 years for men in risk and can be expanded up to 8 years for those not at risk. As for at which age to start screening, options vary between countries and are adapted to the risk of patients [32, 36].

Regarding diagnosis, the current standard procedure is transrectal ultrasound-guided prostate biopsy, following abnormal results obtained from DRE and PSA. By convention, 10-12 tissue samples are collected, and further analyzed histologically for PCa confirmation [36].

2. Liquid biopsy-based cancer biomarkers

A biomarker can be any substance, structure or process measurable in the body that sheds light on disease status and behavior [37]. Particularly, cancer biomarkers comprise a wide variety of molecules such as nucleic acids, enzymes, metabolites, transcription factors, cell surface receptors or epigenetic alterations obtained from either tumor tissue or body fluids [38, 39]. Based on their applications, these can be classified as: “Screening/Diagnostic biomarkers” if they are used to identify early-stage cancer/specific cancer types; “Prognostic biomarkers” if they aim at informing physicians regarding the risk of clinical outcomes; Predictive biomarkers” if they predict response to specific therapeutic interventions [40]. Thus, the ultimate goal of cancer biomarkers is to become a reliable and cost-effective tool for disease management in all its phases, whether from early detection to tumor classification and monitoring, hopefully leading patients to receive the most appropriate treatment [39]. Indeed, with the rise of high-throughput technologies and tumor molecular characterization, biomarkers have become a powerhouse of a field in cancer

research and, together with the demand for precision medicine, these play an increasingly important role in cancer patients' clinical course and, consequently, mortality reduction [38, 40]. Nonetheless, since cancer diagnosis and prognostication use tissue biopsy sampling as a standard approach, biomarker research has also taken advantage of this biological material for molecular profiling of tumors. However, there are important issues linked to tissue-based biopsy that need to be considered: (1) it requires an invasive surgical procedure; (2) some tumors are not accessible due to their anatomical location; (3) its ability to be used as an early detection tool is very reduced, (4) as well as its application in the evaluation of treatment efficiency and monitoring of tumor progression; (5) it does not allow characterization of tumor heterogeneity [41-43].

Recently, liquid biopsies have emerged as a tool to overcome these challenges. Consisting in the capture of tumor-related markers from fluid samples, such as blood or urine, liquid biopsies are an ideal substitute for tissue biopsies due to its minimal invasiveness, ease to obtain, ability to follow-up tumor evolution in real time and highlight tumor heterogeneity [42, 43]. They comprise a variety of analytes namely circulating cell-free DNA (cfDNA), cell-free RNA (cfRNA), circulating tumor cells (CTCs), extracellular vesicles (EVs), tumor-educated platelets (TEPs), proteins and metabolites [43, 44].

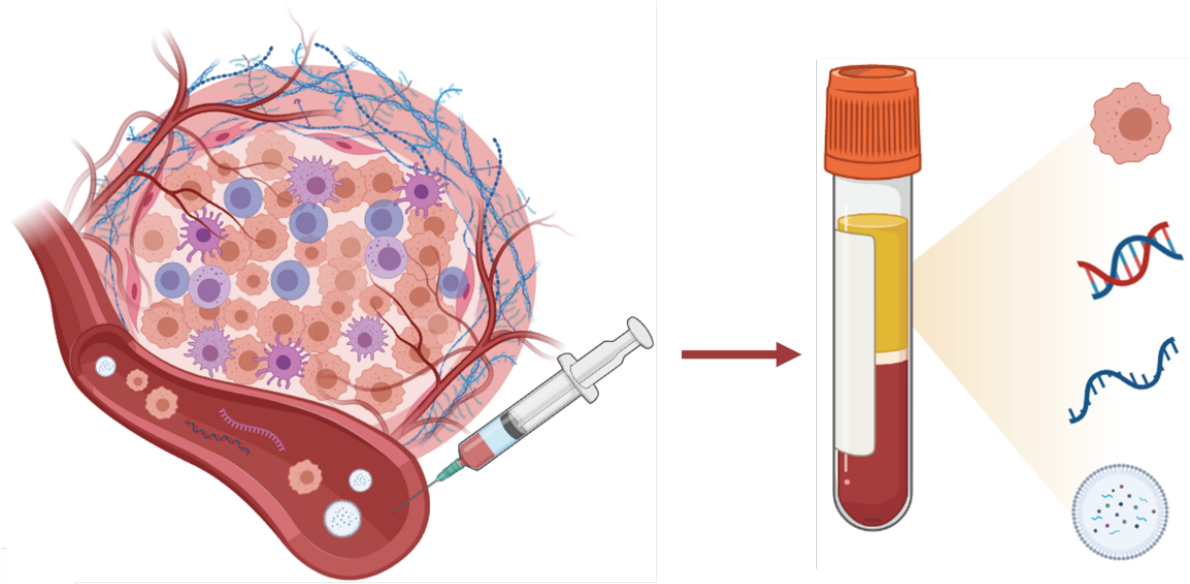

Figure 2. Liquid biopsy. Tumors shed information into the bloodstream, such as circulating cell-free DNA (cfDNA), cell-free RNA (cfRNA), circulating tumor cells (CTCs) and extracellular vesicles (EVs), so blood can be collected and used as a minimally invasive source of cancer-related biomarkers. Created with BioRender.com.

Focusing on cfDNA, consisting in circulating extracellular DNA not related to any subcellular or molecular structure [43], it can arise from a wide variety of cell types and its proportions are also dependent on the physiological status, being the cfDNA of a healthy individual derived primarily from dead blood cells, while a pathological tissue, such as a cancer tissue, can contribute and release more DNA into the circulation [45]. This circulating tumor DNA (ctDNA) can be released into the bloodstream from necrotic or apoptotic cells within the primary tumor, CTCs in the blood, metastatic deposits present in distant locations or it can be even secreted within EVs [46]. Therefore, ctDNA may present the same genetic and epigenetic alterations found in the primary tumor, making it a potential biomarker for cancer detection and monitoring [47]. It is important to note that elevated cfDNA levels in the blood can be due to other conditions besides cancer, such as, inflammation or trauma. Both these events involve cell death, consequently leading to the release of nucleic acids [48]. Hence, ctDNA represents a fraction of cfDNA that can be increased in cancer patients due to the accelerated cellular turnover together with the decreased removal of dead cells, ranging from 0,01% to more than 50% of the total cfDNA population. Other factors that might influence ctDNA quantity are tumor size, location and vascularization, as well as the presence of metastatic sites [47, 48].

Molecular profiling of driver mutations in tumor tissue has been the main clinical use of biomarkers. Accordingly, the first clinical application of liquid biopsies also was the detection of these mutations in ctDNA to replace multiple puncturing with multiple blood draws [49, 50]. In fact, 2 cfDNA-based assays have been approved by the FDA and are currently used to identify patients who will likely respond to specific targeted therapies: cobas® EGFR Mutation Test v2, a PCR test that detects mutations in the *EGFR* gene of patients with NSCLC predictive of response to erlotinib; *therascreen* PIK3CA RGQ PCR kit, a PCR test that detects mutations in the *PIK3CA* gene of BrC patients and predicts response to alpelisib [51]. More recently, 2 next generation sequencing-based assays were also approved: Guardant360 CDx, a genomic profiling test that detects mutations in cfDNA of patients with NSCLC predictive of response to osimertinib and amivantamab [52]; FoundationOne Liquid CDx, the first approved pan-cancer cfDNA-based genomic profiling assay that detects mutations in patients with NSCLC, metastatic PCa, BrC and ovarian cancer, and predicts response to several therapies [53].

Although the use of liquid biopsies for cancer prognostication and drug response prediction has been well established, its utility as a tool for cancer early detection and diagnosis is far behind. However, it has been reported that tumors start to shed information into the circulation very early, allowing to detect cancer even when individuals have not started to develop symptoms or tumor masses are not visible by imaging [54, 55].

3. Epigenetics

The concept of epigenetics arose in the 1940s by Conrad Waddington to explain why genetic variations did not always lead to phenotypic variations and how genes interacted with their environment to yield a phenotype. This concept evolved and, currently, epigenetics is known as the study of gene expression alterations, without changes in the DNA sequence itself. Besides, epigenetic modifications are maintained during cell division, thus being heritable [56, 57]. The main epigenetic mechanisms are DNA methylation, histone tail post-translation modifications, histone variants and chromatin remodelling by protein complexes (Figure 2) [57, 58]. These mechanisms are essential for proper gene expression, defining where and when DNA transcription can start, in order to guide normal cell development and differentiation [59]. Hence, the disruption of epigenetic regulation leads to uncontrolled cell division and differentiation defects, underlying a variety of pathologies, such as cancer [60]. Indeed, aberrant epigenetic patterns are increasingly being studied as alternatives to genetic variations in the disruption of normal gene expression that underlies the carcinogenic process [57].

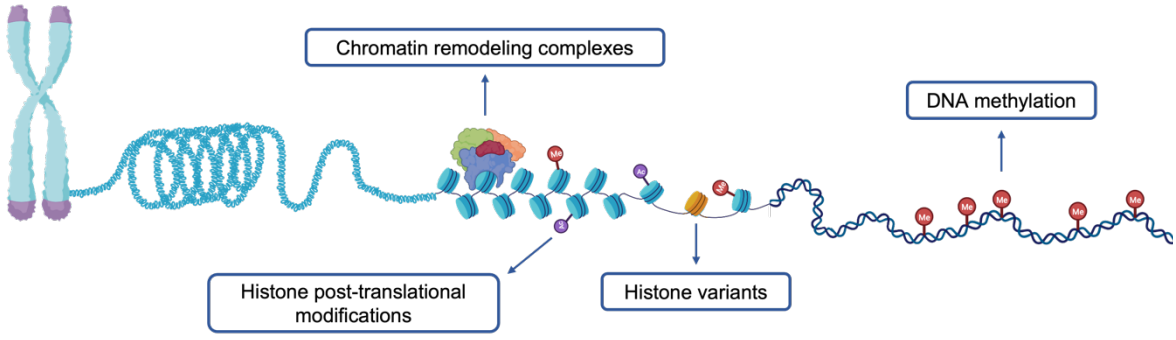

Figure 3. Epigenetic mechanisms regulating gene expression. DNA methylation consists in the addition of a methyl group to cytosines present in CG dinucleotides. Histone post-translation modification consist in the addition of chemical groups (methyl, acetyl, phosphate, ubiquitin, etc.) to amino acid residues of histone tails. Histone variants substitute canonical histones in the nucleosome, impacting chromatin structure. Chromatin remodeling complexes are protein complexes that regulate chromatin structure by altering nucleosome positioning. Created with BioRender.com.

3.1 DNA methylation

DNA methylation is the most well studied epigenetic mechanism in mammals [57-59]. It consists in the covalent addition of a methyl group to the 5-carbon of cytosines residues by the action of DNA methyltransferases (DNMTs). Usually, this reaction takes place within CpG dinucleotides i.e., regions of the genome where a cytosine nucleotide is followed by

a guanine nucleotide. While most CpG dinucleotides are scattered across gene coding regions and repetitive sequences, CpG clusters can be found in the so called CpG islands, preferentially located at the 5' end of the genes, where the promoters are located [61-63].

In normal cells, CpG islands are usually unmethylated, however, they can become methylated, leading to gene silencing. There are some exceptions, and physiological hypermethylation associated with promoters is observed on the silenced copy of the X chromosome in females, on imprinted genes, and in a tissue-specific manner [57]. Gene silencing mediated by DNA methylation can happen by two mechanisms: by directly preventing the interaction of transcription factors with the DNA sequence and/or by the action of methylcytosine-binding proteins (MBPs) that identify methylated CpGs and recruit other proteins, like DNMTs and histone deacetylases (HDACs), that modify the chromatins conformation, preventing gene transcription [57, 63].

Besides, CpG island promoter hypermethylation is commonly linked to cancer, due to altering specific genes' expression, such as tumor suppressor genes or genes involved in DNA repair [61]. In fact, aberrant DNA methylation can contribute to the carcinogenic process by: global hypomethylation of the genome leading to activation of silenced oncogenes; hypermethylation of the promoters of tumor suppressor genes and direct mutagenesis; CpG sites methylation increases the binding of some chemical carcinogens to DNA and increases the rate of UV-induced mutations [62, 64]. Of note, methylation's effect in cells depends on the genomic location where it occurs. As aforementioned, gene promoter methylation is associated with gene silencing, however, gene body methylation appears to be associated with gene expression [65].

The number of known genes affected by transcription disruption through DNA methylation continues to grow and involves genes found at all chromosome locations. Furthermore, this alteration occurs during all tumor progression and, apparently, specific genes are aberrantly methylated in specific tumor stages, which renders DNA methylation an attractive cancer biomarker from early detection to disease monitoring. Additionally, due its cell- and tissue-specificity, combined with its easiness of access through liquid biopsies, DNA methylation presents an ideal biomarker for detecting and discriminating different cancer types in a minimally invasive manner [64, 66]. In fact, cfDNA methylation-based assays have already been developed for cancer detection. For instance, the Epi ProColon® 2.0 test, developed by Epigenomics AG, is intended for CRC detection in individuals declining conventional screening. By assessing the methylation levels of the *SEPT9* gene in cfDNA, this test has shown 75-81% sensitivities and 96-91% specificities for CRC detection and, remarkably, it was the first epigenetic-based blood test for cancer detection to receive FDA approval in 2016 [67, 68]. Additionally, it is also commercially available in Europe (CE-IVD certified) and China [69]. Interestingly, the same company also developed

the Epi proLung® test, meant for LC detection by evaluating the methylation status of the *SHOX2* and *PTGER4* genes. This test displayed 85% sensitivity for 50% specificity, while sensitivity decreased to 59% if 95% specificity was considered [67]. Notably, Epi proLung® received CE-IVD mark in 2017 [67, 68].

4. DNA methylation testing in liquid biopsies as a novel strategy for multi-cancer early detection

Current screening methods for the four major cancers applied in the clinical practice have, indeed, increased the detection and survival rates of cancer patients. Nonetheless, several drawbacks are also associated with each screening method, as mentioned above. While BrC, LC and PCa screening result in high rates of false-positives and overdiagnosis, leading to unnecessary invasive tissue biopsies and treatments, CRC screening is highly invasive. Besides, population-based screening is only available for BrC and CRC, in addition to cervical cancer, meaning that more than 60% of cancer deaths are caused by cancers without any screening option [1]. It is widely accepted that early detection of cancer results in better outcomes for patients, due to the availably of treatment options, namely curative ones. In fact, at least 15% of cancer-related deaths within 5 years could be avoided by detecting the disease early onset [70]. Thus, there is a demand for new screening and early detection strategies, preferably an unexpensive, minimally invasive and highly sensitive and specific one. Following such rationale, a blood-based test that could simultaneously detect multiple cancer types in early stages, and even be applied to high-risk population-based screening, seems a rather intriguing new approach. Besides, a pan-cancer approach might be the only cost-effective option for low prevalent cancer screening [71]. Such a multi-cancer early detection (MCED) test would ideally have high sensitivity for early-stage disease detection, high specificity to avoid false-positive results, and the ability to discriminate the tissue of origin (TOO) of the detected cancer [71]. Besides, given DNA methylation's features of aberrant tumor-specific patterns, tissue-specificity and easiness to assess in cfDNA, there is great potential for methylation-based pan-cancer detection [66].

Having this in mind, we conducted a literature review aiming to explore the current approaches and advances being made for multi-cancer early detection using DNA methylation as a biomarker (Appendix I). The information gathered from the studies found and considered relevant is displayed in Table 1, showing multi-cancer detection strategies validated in human clinical specimen.

Table 1. DNA methylation-based multi-cancer early detection (MCED) tests.

Biomarker	Source	Tumor types	Sample	Methodology	Main findings	Ref.
DNA methylation	Tissue	lung, breast, colorectal, esophagus, liver, pancreatic, gastric, cervical, head and neck	120 tumor tissue 123 normal tissue	Bisulfite pyrosequencing	TCGA methylation data mining identified *HIST1H4F* as hypermethylated in 17 tumor types. Methylation analysis in tissue samples of 9 cancer types showed AUCs above 0.87 for all cancers and above 0.90 for all except pancreatic cancer.	[72]
	Tissue	lung, breast, colorectal, prostate, pancreas, glioblastoma and leukemia	83 tumor tissue 54 normal tissue	Bisulfite pyrosequencing	Methylation levels at 27 CpGs of the *GHSR* gene showed a higher average methylation degree in all tumor samples compared to normal samples. 27 CpG-signature displayed an AUC of 0.8789 for discriminating cancer from normal tissue.	[73]
	Tissue	colorectal, gastric and esophageal	229 tumor and normal-adjacent tissue	Bisulfite-sequencing PCR	TCGA methylation data mining identified differentially methylated regions (DMRs) in the *SST* gene. 7 CpG sites were shown to be hypermethylated in all 3 cancers. A combination of 2 CpGs (+18 and +129) displayed the best AUC of 0.698, with 59.3% sensitivity and 72.8% specificity for detecting the 3 gastrointestinal cancers.	[74]
	Tissue	lung, breast, colon, gastric and endometrial	184 tumor tissue 34 normal tissue	Bisulfite amplicon sequencing	Designed a 302-bp PCR amplicon, covering the *ZNF154* tumor-specific hypermethylated region, and methylation patterns were used to develop a multi-cancer classifier. AUC of 0.96 for discriminating cancer from normal tissue. Computational simulation of ctDNA displayed AUCs of up to 0.79.	[75]
	Plasma	colon, pancreatic, liver and ovarian	71 cancer patients 20 healthy individuals	DREAMing	TCGA methylation data from white blood cells revelead that *ZNF154* locus remains unmethylated, even in older individuals, showing the potential for the development of a blood test for cancer detection. AUC values ranged from 0.75 to 0.87 for discriminating cancer patients from healthy individuals, except for liver cancer which displayed an AUC of 0.48.	[76]
	Plasma	lung and prostate	323 cancer patients 136 healthy individuals	qMSP	"PanCancer" panel (*FOXA1, RARβ2* and *RASSF1A*) detected cancer with 64.3% sensitivity, 69.8% specificity and 66.4% accuracy. "CancerType" panel (*GSTP1* and *SOX17*) discriminated between lung and prostate cancer with 93% specificity.	[77]
	Plasma	lung, breast and colorectal	253 cancer patients 103 healthy individuals	qMSP	"PanCancer" panel (*APC*, *FOXA1*, *RASSF1A*) detected cancer with 72.4% sensitivity, 73.5% specificity and 72.8% accuracy. "CancerType" panel (*SCGB3A1, SEPT9* and *SOX17*) discriminated TOO with 80.0%, 98.9% and 85.1% specificity for breast, colorectal and lung cancer, respectively.	[78]
	Serum	lung, breast, colorectal, gastric, pancreatic, and hepatocellular	70 cancer patients 10 healthy individuals	MSP	Methylation levels of a 4 gene-panel (*RUNX3, p16, RASSF1A* and *CDH1*) showed 89% sensitivity and 100% specificity for cancer detection.	[79]
	Plasma	colorectal and pancreatic	60 cancer patients 60 healthy individuals	Methylation array	Found a 7 gene panel (*MDR1, SRBC, VHL, MUC2, RB1, SYK* and *GPC3*) that detects colorectal and pancreatic cancers with 63.16% sensitivity, 84% specificity and AUC of 0.8177.	[80]
	Plasma	lung, breast and liver	46 cancer patients 32 healthy individuals	Bisulfite sequencing	Developed CancerLocator, a test based on cfDNA bisulfite sequencing combined with a probabilistic model for cancer detection and TOO discrimination. CancerLocator uses TCGA methylation data as features to estimate the fraction of ctDNA in the plasma and the likelihood of coming from each tumor type. TOO discrimination showed a low error rate of 0.265 (99.7% accuracy).	[81]

	Plasma	liver but applicable to any cancer	33 cancer patients 36 healthy individuals	Bisulfite sequencing	Developed CancerDetector, a test based on cfDNA bisulfite sequencing combined with a probabilistic model that joints methylation states of multiple adjacent CpG sites on an individual sequencing read, for cancer detection. 94.8% sensitivity and 100% specificity were obtained.	[82]
	Plasma	> 50 cancer types	2482 cancer patients 4207 healthy individuals	Bisulfite sequencing	Developed a targeted methylation assay combined with a machine learning classifier for detecting and discriminating TOO in more than 50 cancer types using cfDNA. 54.9% sensitivity and 99.3% specificity were obtained in the validation set. 93% accuracy for TOO prediction.	[83]
			2 823 cancer patients 1 254 healthy individuals		Developed a refined assay and classifiers optimized for screening purposes and performed clinical validation. 51.5% sensitivity and 99.5% specificity were obtained. 88.7% accuracy for TOO prediction. PPV of 44.4% and NPV of 99.4% for cancer detection.	[84]
	Plasma	colorectal, hepatocellular, esophageal, gastric and pancreatic	254 cancer patients 46 healthy individuals	Bisulfite sequencing	Developed EpiPanGI Dx, a cfDNA methylation-based test combining bisulfite sequencing and machine learning, for detecting and discriminating TOO of gastrointestinal cancers. AUC of 0.88 for detecting gastrointestinal cancers. Accuracy of 0.85-0.95 for TOO prediction.	[85]
	Plasma	lung, colorectal, gastric, liver and esophageal	191 pre-diagnosis cancer samples 223 post-diagnosis cancer samples 414 healthy samples	Bisulfite sequencing (using semi-targeted PCR libraries)	Developed PanSeer, a blood test combining the analysis of 477 cancer-specific differentially methylated regions with machine learning for cancer detection. 87.6% sensitivity for post-diagnosis samples, 94.9% sensitivity for pre-diagnosis samples and 96.1% specificity were obtained in the testing set. Cancer can be detected by PanSeer up to 4 years before conventional diagnosis with 95.7% sensitivity.	[86]
	Plasma	lung, pancreatic and acute myeloid leukemia	137 cancer patients 62 healthy individuals	cfMeDIP-seq	Developed cfMeDIP-seq, an immunoprecipitation-based protocol for methylation profiling in cfDNA and combined it with machine learning algorithms to discriminate TOO. AUC values ranged from 0.92 to 0.98 for discriminating TOO.	[87]
	Plasma	lung, breast, colorectal and melanoma	78 cancer patients 66 healthy individuals	Bisulfite sequencing	Developed a targeted methylation sequencing assay to analyze the methylation status of 9 223 cancer related CpG sites, combined with a novel algorithm that converts sequencing data into a methylation score, for cancer detection and TOO discrimination. 83.8% sensitivity and 100% specificity were obtained for cancer detection. 78.9% accuracy for TOO discrimination.	[88]
	Plasma	lung, breast, colorectal and liver	Not available	NGS	Developed IvyGeneCORE Test, a blood test analyzing cfDNA methylation levels at specific genes combined with artificial intelligence for cancer detection. 84% sensitivity and 90% specificity were obtained for discriminating cancer from healthy individuals.	[89]
	Plasma	lung, colorectal, pancreatic, liver, esophageal and ovarian	625 cancer patients 483 healthy individuals	ELSA-seq	Developed ELSA-seq, a targeted methylation sequencing assay combined with machine learning for cancer detection and TOO discrimination. 80.6% sensitivity and 98.3% specificity were obtained in validation set. 81.0% accuracy for TOO discrimination.	[90]
	Plasma	14 cancer types	549 cancer patients 80 healthy individuals	Targeted sequencing	Developed a cancer detection model based on 37 methylation-correlated blocks (MCB). 72.86% sensitivity, 96.67% specificity and AUC of 0.86 were obtained in the validation set.	[91]
	Plasma	lung, breast, colorectal, pancreatic, gastric, esophageal, liver and ovarian	598 cancer patients 302 healthy individuals	Targeted sequencing	Developed a cancer detection and TOO discrimination model based on 135 MCB. 66.3% sensitivity, 95.5% specificity and AUC of 0.85 were obtained in the validation set. 75.4% accuracy for TOO discrimination.	[92]

	Plasma	lung, breast, colorectal and pancreatic	101 cancer patients 71 healthy individuals	MSRE-qPCR	Developed a 10-marker panel for cancer detection and a 16-marker panel for TOO discrimination. 79% sensitivity, 90% specificity and AUC of 0.89 were obtained for cancer detection. TOO discrimination accuracy was 80% for colorectal, 78% for lung, 75% for pancreatic and 62% for breast cancer.	[93]
	Plasma	lung, colorectal, bladder and pancreatic	> 1 500 cancer patients > 1 800 healthy individuals	5mC enrichment and targeted sequencing	Developed a blood test based on cfDNA methylation signatures for early cancer detection and TOO discrimination. 90% and 87% sensitivity at 90% specificity for stage I/II colorectal and lung cancer detection. 73% and 52% sensitivities at 95% specificity for stage I/II pancreatic and bladder cancer detection. At 98% specificity, TOO accuracy was 99% for colorectal, 94% for lung, 88% for bladder and 86% for pancreatic cancer.	[94]
	Plasma	lung, breast, colorectal, prostate, pancreatic, liver and ovarian	111 cancer patients 55 healthy individuals	Targeted sequencing	Developed Omni1, a targeted methylation sequencing panel comprising around 3000 cancer-specific hypermethylation markers for cancer early detection. 65% sensitivity for stage I cancers, 75% sensitivity for stage II cancers and 89% specificity were obtained.	[95]
	Plasma	lung, breast, colorectal, gastric, esophageal and liver	269 cancer patients 170 healthy individuals	Bisulfite sequencing	Developed Aurora, a blood test based on cancer specific cfDNA methylation signatures for detecting 6 major cancer types. AUCs of 0.90, 0.98 and 0.92 were obtained for lung, breast and colorectal cancer detection, respectively.	[96]
			203 cancer patients 206 healthy individuals		Improved to Aurora 2.0, a targeted methylation sequencing assay. AUCs of 0.94 and 0.935 were obtained for gastric and esophageal cancer detection, respectively. AUCs of 0.973, 0.962 and 0.92 were obtained for lung, breast and colorectal cancer detection, respectively.	[97]
			1000 cancer patients 505 healthy individuals		AUCs of 0.973, 0.962 and 0.92, 0.94 and 0.935 were obtained for lung, breast, colorectal, gastric and esophageal cancer detection, respectively. At 99% specificity, 84%, 75%, 82%, 85% and 78% sensitivity were obtained for lung, breast, colorectal, gastric and esophageal cancer, respectively.	[98]
	Tissue Plasma	breast, colorectal, prostate and lymphoma	72 tumor and 31 normal tissues 100 cancer and 45 healthy plasmas	Electrochemical assays	Developed electrochemical and colorimetric assays that can detect methylation differences between cancer and healthy genomes based on the level of DNA adsorption on planar and colloidal gold surfaces. DNA adsorption levels could discriminate between cancer patients and healthy individuals with an AUC of 0.887 using an electrochemical assay. DNA adsorption levels could discriminate between cancer patients and healthy individuals with an AUC of 0.785 using a colorimetric assay.	[99]
	Stool	colorectal and gastric	105 cancer patients 113 healthy individuals	Hi-SA	Developed a method combining single-step sodium bisulfite modification and fluorescence PCR to measure RASSF2 and SFRP2 methylation status in fecal DNA. DNA recovery from feces showed an AUC of 0.78 for distinguishing cancer from non-advanced lesions (adenomas, polyps and healthy). Methylation levels showed an AUC of 0.78. A combination score showed the best AUC of 0.81.	[100]
DNA methylation and circulating proteins	Plasma Serum	lung, pancreatic, gastric, esophageal, liver and ovarian	180 cancer patients 257 healthy individuals	Multiplex PCR + LQAS	Developed a multi-analyte blood test based on 26 methylation markers and 5 circulating proteins combined machine learning algorithms for cancer detection. 83% sensitivity, 94% specificity and AUC of 0.96 were obtained in validation set.	[101]
			160 cancer patients 315 healthy individuals		85% sensitivity, 95% specificity and AUC of 0.96 were obtained in validation set.	[102]

DNA methylation and copy number variations	Plasma	lung, breast, hepatocellular, nasopharyngeal, smooth muscle sarcoma and neuroendocrine tumor	46 cancer patients 32 healthy individuals	Bisulfite sequencing	Performed bisulfite sequencing to analyze genome-wide hypomethylation combined with copy number alterations in cfDNA and developed algorithms for cancer detection. If a sample was positive if either hypomethylation or CNAs were observed, 85% sensitivity and 88% specificity were obtained. If a sample was positive if both hypomethylation and CNAs were observed, 60% sensitivity and 94% specificity were obtained.	[103]
cfDNA methylation, fragmentation, copy number variations and microbial composition	Plasma	lung, colon, gastric and liver	275 cancer patients 204 healthy individuals	cfMethyl-Seq	Developed CancerRadar, a test based on genome-wide methylation profiling of cfDNA combined with machine learning for cancer detection and TOO discrimination. 85.6% sensitivity and 99% specificity for cancer detection. 91.5% accuracy for TOO discrimination.	[104]
DNA hydroxymethylation	Plasma	lung, breast, colorectal, gastric, esophageal and liver	2 241 cancer patients 2 289 healthy individuals	5hmC-Seal profiling	Used the 5hmC-Seal technology to profile genome-wide 5hmC in cfDNA and combined it with machine learning for cancer detection and TOO discrimination. 79.3% sensitivity and 95% specificity were obtained in training set. 67.6% sensitivity and 98.2% specificity were obtained in the testing set. 83.2% accuracy for TOO discrimination.	[105]
	Plasma	lung, breast, prostate and pancreatic	188 cancer patients 180 healthy individuals	5hmC enrichment and sequencing	Developed a novel 5hmC enrichment technology coupled with sequencing and machine learning for cancer detection. AUCs of 0.89, 0.84, 0.95 and 0.83 were obtained for breast, lung, pancreatic and prostate cancer detection, respectively.	[106]

Abbreviations: AUC – Area under ROC curve; cfDNA – cell-free DNA; cfMeDIP-seq – cell-free methylated DNA immunoprecipitation and sequencing; CpG – Cytosine-phospate-Guanine; ctDNA – circulating tumor DNA; DREAMing - Discrimination of Rare EpiAlleles by Melt; Hi-SA – high-sensitivity assay for bisulfite DNA; LQAS - long probe quantitative amplified signal; MSP – methylation-specific PCR; MSRE-qPCR – methylation-sensitive restriction enzyme -based quantitative PCR; NGS – next generation sequencing; NPV – negative predictive value; PPV – positive predictive value; qMSP – quantitative methylation-specific PCR; TCGA – The Cancer Genome Atlas; TOO – tissue of origin; 5hmC – 5-hydroxymethylcytosine; 5mC – 5-methylcytosine.

Whether analyzing a single gene [76] or gene panels [77, 78], methylation levels of cfDNA have demonstrated the feasibility of minimally invasively detecting multiple cancers and further identifying their anatomical location. Nonetheless, these approaches fall short regarding sensitivity values. On the other hand, sequencing-based methylation profiling of cfDNA has shown more promising results, by using machine learning algorithms that convert the complex data into classifiers that discriminate cancer from healthy individuals and further identify its origin. For instance, Kandimalla et al. reported EpiPanGI Dx, an assay that showed an AUC of 0.88 for simultaneous detection of gastrointestinal cancers and accuracies of 85-95% for TOO prediction [85]. Focusing on 4 major cancers (lung, breast, colorectal and liver), the IvyGeneCORE® Test is commercialized by the Laboratory for Advanced Medicine, showing that the methylation analysis of target genes discovered by data mining can detect these cancers with 84% sensitivity and 90% specificity [89, 107]. Similarly, the PanSeer assay developed by Singlera Genomics [108] uses semi-targeted PCR libraries followed by sequencing for analyzing 477 differentially methylated regions (DMRs). This blood test was evaluated using samples from the Taizhou Longitudinal Study, in which healthy individuals provided plasma samples and were monitored for cancer, allowing a retrospective take on early detection viability. Concerning 5 tumor types (lung, colorectal, gastric, liver and esophageal), 87.6% sensitivity and 96.1% specificity were obtained, with similar sensitivities between early- and late-stage disease. Remarkably, using pre-diagnostic samples, PanSeer showed that cancer could be detected up to 4 years before medical diagnosis with 95.7% sensitivity [86]. Nevertheless, no results regarding TOO prediction were reported.

A company that revolutionized the cancer screening paradigm and emphasized the wide variety of cancers that can be simultaneously detected through liquid biopsy is GRAIL, a spin-off of Illumina, that received around $1 billion in funding for the solen goal of developing a blood test for early cancer detection [109, 110]. For such purpose, the Circulating Cell-free Genome Atlas Study (CCGA) (NCT02889978), divided into 3 sub-studies, was conducted and recruited over 15000 participants with and without cancer that were longitudinally followed. In the first CCGA sub-study, 3 different sequencing assays were evaluated and, ultimately, whole-genome bisulfite sequencing outperformed whole-genome sequencing and targeted mutation analysis, showing, once more, the superior ability of DNA methylation for early cancer detection [111, 112]. Therefore, in the second sub-study, a targeted methylation assay was developed, trained, and validated using 6689 participants, for simultaneous detection and TOO discrimination of more than 50 cancer types. 54.9% sensitivity and 99.3% specificity were obtained for all cancer stages, while sensitivity was 43.9% for early stages. Besides, when focusing on a set of 12 high-signal cancers (based on Surveillance, Epidemiology, and End Results (SEER) mortality data) sensitivity was

67.3%. Notably, 93% accuracy was displayed for TOO localization [83]. In the third and final sub-study, meant to further validate an improved test version specific for screening purposes, an independent validation set of 5309 participants was used and resulted in 51.5% sensitivity, 99.5% specificity and 88.7% accuracy for TOO prediction [84]. Given the prospective nature of CCGA, the prognostic value of this blood test could also be assessed. By following cancer patients from the second sub-study for 3 years, it was possible to observe that cancers not detected by the test had significantly better overall survival (OS) than those detected by the MCED test. Additionally, detection sensitivity was higher in participants who died than in those alive, suggesting that this test tends to detect more aggressive cancers and may not increase risk of overdiagnosis [113]. Currently, this blood test is available to the population as Galleri® at the price of $949, upon request to health care providers [114]. In addition to CCGA, other clinical trials are being conducted by GRAIL to ripen the tests' potential as a screening tool: STRIVE (NCT03085888) is evaluating the test performance to detect breast and other invasive cancers in women undergoing screening mammography; SUMMIT (NCT03934866) is evaluating the test performance to detect invasive cancers in individuals at high risk of lung and other cancers due to a significant smoking history; PATHFINDER (NCT04241796, NCT05155605) is assessing the implementation of the test into clinical practice; REFLECTION (NCT05205967) wants to understand the performance of the test in clinical settings and its impact on patients and healthcare professionals. In fact, some results from PATHFINDER have already been reported. Aiming to evaluate the time and number of additional procedures required to achieve a final diagnosis following a positive test result, it was observed that a cancer signal was detected in 1.5% of participants, of which 65% reached a diagnostic resolution. The median time for diagnosis was 78 days, with 93% of participants undergoing imaging tests and 72% being submitted to an invasive procedure. Remarkably, only 18% of participants with a final non-cancer diagnosis had to go through an invasive procedure [115, 116].

Most PCR- and sequencing-based methods for methylation analysis rely on sodium-bisulfite modification and it has been proven that this aggressive treatment causes DNA degradation and fragmentation, hindering the analysis of large CpG islands, especially in cfDNA which is already highly fragmented [117]. As an alternative, immunoprecipitation of methylated DNA (MeDIP), i.e., the use of antibodies that target 5-methylcytosine (5mC) for the enrichment of methylated DNA fragments, followed by sequencing can be used [118]. Following such reasoning, Adela is a company developing a sensitive technology for the enrichment of methylated fragments from low input samples, like cfDNA, followed by sequencing of cancer-related regions (cfMeDIP-seq) [119, 120]. When applied to cancer detection, by combining the above-described assay with machine learning, AUC values of 0.980, 0.918, 0.971, and 0.969 were obtained for discriminating acute myeloid leukemia,

pancreatic cancer, lung cancer and healthy individuals, respectively. Besides, early- and late-stage cancer detection depicted similar values [87]. Interestingly, the CAMPERR study (NCT05366881) was, at the time of writing, recruiting patients with any of 20 tumor types, plus healthy individuals to validate the cfMeDIP-seq assay.

To the best of our knowledge, several other methylation-based MCED tests using a variety of methodologies are being currently developed by different companies (Table 1). Besides, many of them are also conducting clinical trials for participant recruitment and prospective assessment of test performance.

Remarkably, methylation analysis showed potential for cancer detection even beyond its molecular analysis. Aberrant DNA methylation patterns in cancer also alter the physicochemical properties of DNA, which led Sina et al. to develop simple, fast analysis and low-input electrochemical and colorimetric assays. These demonstrated AUC values of 0.887 and 0.785, respectively, for differentiating breast and colorectal cancer from control plasma samples [99]. However, only advanced-stage samples were used, suggesting that, although promising, these prototypes require validation in early-stage cancer as well as more tumor types.

Beyond 5mC modifications, 5-hydroxymethylcytosine (5hmC), the result of 5mC oxidation catalyzed by Ten-Eleven Translocation (TET) enzymes [121], was also proposed as pan-cancer biomarker by Li et al., who developed a genome-wide 5hmC analysis tool and reported 67.6% sensitivity and 98.2% specificity for cancer detection and 83.2% accuracy for TOO discrimination in 6 cancer types [105]. Additionally, BlueStar Genomics is also conducting a study (NCT03869814) for the development of a 5hmC-based MCED test and already reported some promising results [106, 122].

Interestingly, approaches, other than methylation, to multi-cancer early detection are also being developed. As early as 2009, Zou et al. performed targeted mutation analysis in stool from several gastrointestinal cancer patients and showed that pan-gastrointestinal cancer detection was feasible with 68% sensitivity and 100% specificity [123]. In fact, stool is also a non-invasive source of cancer biomarkers but limited to tumors of the digestive system. Curiously, a study evaluating patients' perceptions about stool-based multi-cancer detection reported that 98% of participants would use such a test, preferred it over conventional colorectal cancer screening, and highlighted its pan-cancer feature as the most important [124]. Cohen et al. also reported a blood test, CancerSEEK, for the detection of 8 common cancers (lung, breast, colorectal, pancreatic, gastric, liver, esophageal, and ovarian) based on the analysis of mutations in 16 genes combined with the circulating levels of 8 proteins. Methodologically, this test consists of a multiplex PCR and a single immunoassay, being a simple workflow and easily applicable to clinical practice, with an estimated price of around $500. When applied to 1005 cancer patients and 812 healthy

controls, CancerSEEK showed 62% sensitivity at a specificity greater than 99% for discriminating cancer from healthy samples. Concerning early-stage detection, a median sensitivity of 43% was obtained for stage I, 73% for stage II, and 78% for stage III. Additionally, TOO discrimination was also possible with 63% accuracy [125]. However, it is important to note that the protein biomarkers were the major contributors to identifying the underlying cancer type after a positive test result. A refined version of CancerSEEK was then developed and combined with PET-CT imaging to evaluate the test performance for detecting cancer prospectively in a study (DETECT-A) involving 10,006 women not known to have cancer. For that purpose, participants were submitted to the blood test, if results were abnormal a second blood collection was conducted for confirmation and, if such came back positive, a full body PET-CT was performed. Test results were considered positive for 134 participants, out of which 127 were further evaluated by PET. 64 showed concerning imaging and 26 were proven to have cancer. This resulted in 27.1% sensitivity and 98.9% specificity for blood testing alone, while sensitivity decreased to 15.6% and specificity increased to 99.6% for blood testing combined with PET imaging [126]. Therefore, although mutation-based MCED tests have demonstrated their potential, even in early stages, these might not be the ideal approach, since TOO identification of the detected cancers is not possible, due to gene driver mutations not being tissue-specific [127].

Since the mechanisms of cell death causing DNA shedding into the blood are variable, reflecting different fragmentation patterns, and also are cell- and tissue-dependent mechanisms, reflecting nucleosome positioning in the nucleus, tumor-derived cfDNA fragments carry distinct features that may allow cancer detection and further TOO identification [127, 128]. Indeed, DELFI Diagnostics developed the DELFI assay, which, using genome-wide fragmentation analysis in 236 cancer patients (lung, breast, colorectal, pancreatic, gastric, bile duct and ovarian) and 245 healthy individuals, displayed 73% sensitivity and 98% specificity for discriminating cancer from healthy, and 61% accuracy for TOO [129, 130]. Noteworthy, when combining mutation analysis with fragmentation, DELFI showed an increased sensitivity and TOO accuracy of 91% and 75%, respectively [130]. This suggest that combining cfDNA fragmentation with other tumor-related feature may increase its detection capacity. Interestingly, CancerRadar, a multi-omics approach combining cfDNA fragmentation with methylation, copy number variations and microbial composition showed a remarkable 85.6% sensitivity and 99% specificity for lung, colon, gastric and liver cancer detection and 91.5% TOO accuracy [104].

Overall, these data suggest the feasibility of using a single blood test for simultaneously detecting several cancer types, with DNA methylation-based tests being in the forefront of development. Hopefully, such tests will soon be available to the general public to increase patient compliance to screening programs and reduce cancer-related mortality.

II.PRELIMINARY RESULTS

This study is integrated in a larger project developed in the Cancer Biology and Epigenetics Group (CI-IPOP), whose major goal is to develop a “PanCancer” and “CancerType” gene panels for early detection of the four major cancers, using cfDNA methylation as a biomarker (CI-IPOP-74-2016-MethylBiom4Can).

The “PanCancer” approach aims to select the most sensitive and specific gene panel for simultaneous detection of the four cancers, while a “CancerType” gene panel focuses on selecting the most adequate genes for discriminating each individual cancer.

Some results gathered until the beginning of this project include:

TCGA data mining for gene selection

TCGA methylation data was retrieved from the Shiny Methylation Analysis Resource Tool (SMART) App website, an interactive web server for analysing DNA methylation of TCGA project [131]. Differential Methylation Analysis was performed on TCGA-LUSC & TCGA-LUAD (lung cancer), TCGA-BRCA (breast cancer), TCGA-COAD (colon cancer) and TCGA-PRAD (prostate cancer) datasets using the following criteria: Hypermethylation, M-Value, M-value cut-off = 2 and Adj p.value cut-off = 0.01. This resulted in the identification of 15,833 CpG probes as hypermethylated in TCGA-LUSC dataset, 7,589 CpGs in TCGA-LUAD, 10,550 CpGs in TCGA-BRCA, 16,122 CpGs in TCGA-COAD and 9,976 CpGs in TCGA-PRAD (Figure 4).

A Venn diagram (Figure 5) was then assembled to evaluate the number of CpG probes that were hypermethylated in all four cancer types (PanCancer) and only in each one of the cancer types (CancerType). 494 CpGs were identified for “PanCancer” panel, whereas 796 CpGs were identified exclusively for LC, 3491 CpGs for BrC, 9399 CpGs for CRC and 4202 CpGs for PCa.

Later, the data obtained was merged with data from HumanMethylation450 v1.2. Manifest File (Illumina) and CpG probes were filtered by UCSC CpG Island (Island) and USCS RefGene_group (TSS200 or TSS1500) to obtain the number of CpGs located concomitantly in CpG Islands and promoter regions. Finally, as each gene might contain more than one hypermethylated CpG, the number of genes represented by more than one CpG island were counted (Figure 4). Therefore, 30 genes were identified as promising genes for our “PanCancer” panel, whereas 12 genes for LC discrimination, 82 for BrC, 496 for CRC and 159 for PCa. The top 10 hypermethylated genes for PanCancer and top 5 for each cancer type are described in Table 2.

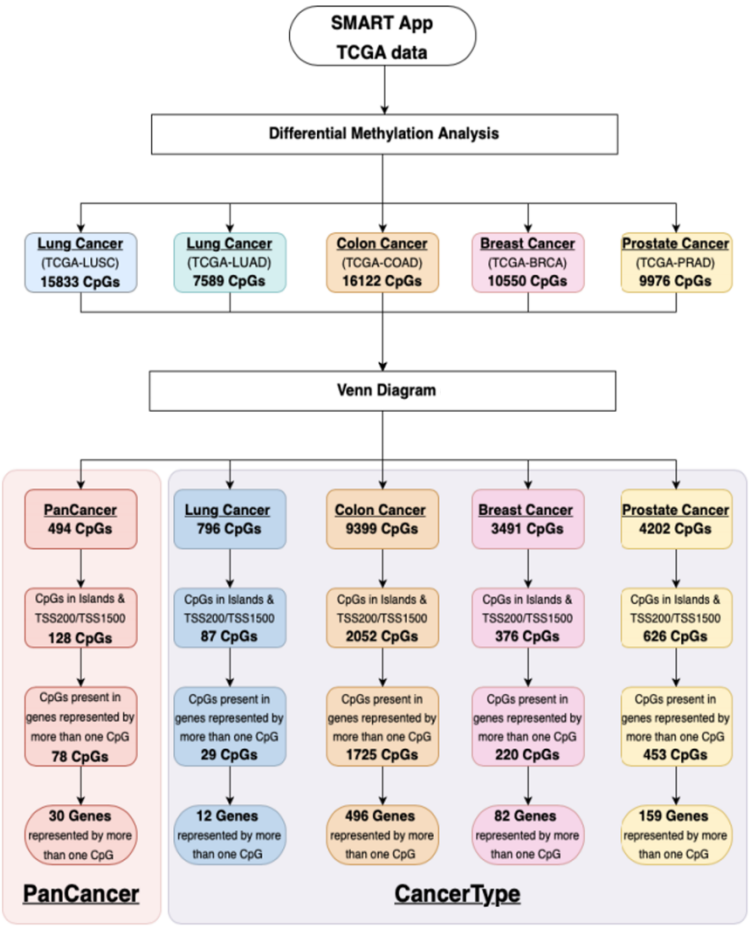

Figure 4. Flowchart of TCGA data mining analysis performed. Kindly provided by V. Constâncio. Unpublished.

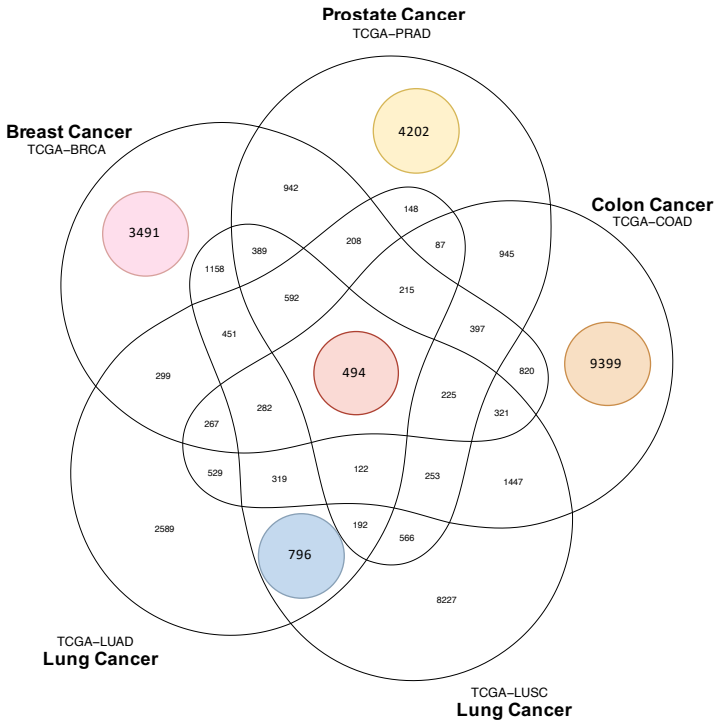

Figure 5. Venn Diagram obtained from TCGA data mining regarding the number of CpG probes listed as hypermethylated in TCGA datasets for each cancer type. Kindly provided by V. Constâncio. Unpublished.

Table 2. Top hypermethylated genes for PanCancer (Top 10) and for each cancer type (top 5) depicted from TCGA data mining analysis. Bold genes indicate the genes selected for further analysis in this project.

Cancer Type	Gene	Number of hypermethylated CpGs
PanCancer		
	ADCY4	5
	CDO1	4
	MAGI2	4
	MIR129-2	4
	NID2	4
	DRD5	3
	EMX2	3
	OLIG2	3
	PAX6	3
	POU3F3	3
Lung Cancer		
	HOXA11	5
	ZNF529	3
	ZSCAN11	3
	BARX2	2
	CETN1	2
Breast Cancer		
	CELF2	7
	EDNRB	6
	FOXA2	6
	GFI1	6
	ISM1	5
Colon Cancer		
	CHFR	19
	GFRA1	18
	BRUNOL4	14
	HSPA1L	14
	NDRG4	12
Prostate Cancer		
	FLOT1	18
	HRASLS5	7
	AOX1	6
	C2orf88	6
	CCND2	6

In silico analysis

In silico analysis was performed to evaluate the ability of the selected genes to distinguish tumoral from normal tissue for each cancer type. For this purpose, mean aggregation methylation data of the most relevant CpGs, depicted in the data mining, for each gene was retrieved from the SMART App website (Figure 6). Mann-Whitney non-parametric test was used for comparisons between the two groups.

As expected, *ADCY4*, *CDO1*, *MAGI2*, *MIR129-2* and *NID2* aggregation methylation levels were significantly higher in tumor tissues of the four cancer types when compared to normal tissues ($p<0.0001$ in all comparisons) (Figure 6 A, B, C, D, E). These results are in agreement with the obtained data mining results (Figure 4 and 5; Table 2), thereby, being potential candidates for a "PanCancer" gene panel, aiming to simultaneously detect these four cancers. Regarding *HOXA11*, *CELF2*, *CHFR* and *FLOT1* aggregation methylation levels, each gene promoter exhibited higher levels for lung, breast, colorectal and prostate tumor tissues, respectively, when compared to normal tissues ($p<0.0001$ in all comparisons). Although significant differences were also observed for these genes between other tumor types and respective normal tissues, we could observe that β-values' amplitude was higher for the specific cancer type when compared to others (Figure 6 F, G, H, I). Hence, *HOXA11*, *CELF2*, *CHFR* and *FLOT1* seem to be potential markers for discrimination of LC, BrC, CRC and PCa, respectively.

Validation of selected genes in tissue samples

Methylation-specific primers and probes were designed in the promoter region of the most promising hypermethylated genes depicted in the data mining analysis. The promoter methylation levels of the 8 most promising genes (*CDO1* was not analyzed due to lack of specificity of the primers/probe designed) were then tested in tissue samples of 120 tumors and 61 controls from IPO-Porto patients by quantitative methylation-specific PCR (qMSP).

ADCY4, MAGI2, MIR129-2, NID2 promoter methylation levels were significantly higher in tumor tissues when compared to normal tissues independently of the tissue of origin (Figure 7 A, B, C, D), suggesting their putative value as "PanCancer" biomarkers for liquid biopsy testing. Contrarily, *FLOT1* was observed to be hypermethylated in PCa samples (Figure 7H), whereas *CELF2* and *CHFR* were hypermethylated in BrCa and CRC (Figure 7F-G), respectively, but in a lesser extent. Disappointingly, although significant differences were only depicted between LCa and NL (Figure 7E), high *HOXA11* methylation levels were depicted in LCa, BrCa, NBr and some CRC and PCa, indicating the lack of specificity as a "CancerType" biomarker.

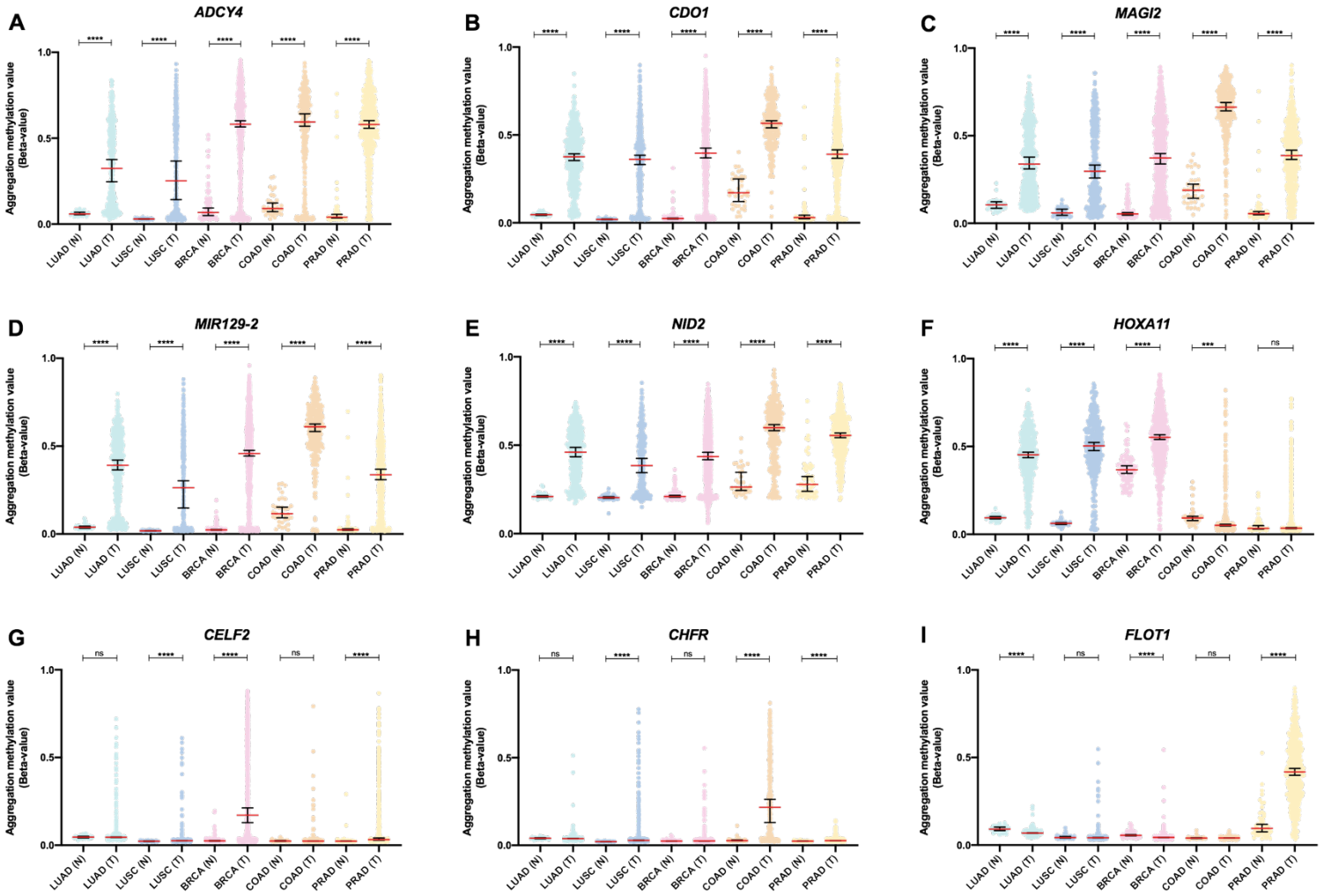

Figure 6. Distribution of (A) *ADCY4*, (B) *CDO1*, (C) *MAGI2*, (D) *MIR129-2*, (E) *NID2*, (F) *HOXA11*, (G) *CELF2*, (H) *CHFR* and (I) *FLOT1* aggregation methylation levels of selected CpGs in lung cancer adenocarcinoma (LUAD), lung cancer squamous cell carcinoma (LUSC), breast cancer (BRCA), colorectal cancer (COAD) and prostate cancer (PRAD) tumor (T) and normal tissues (N). Mann-Whitney U Test between tumor and normal tissues, n.s. $p>0.05$, *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$. Red horizontal lines represent the median methylation level.

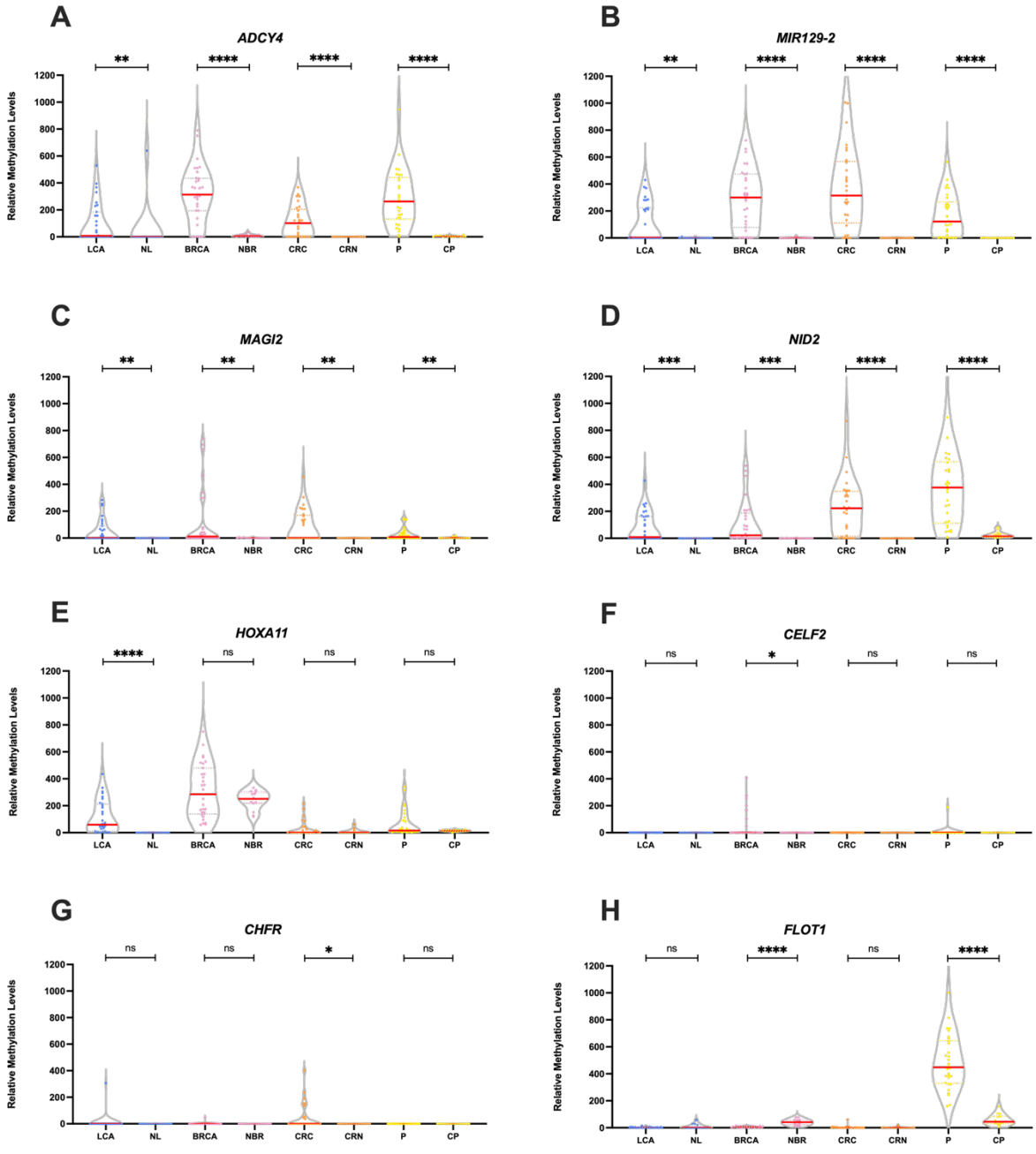

Figure 7. Distribution of (A) *ADCY4*, (B) *MAGI2*, (C) *MIR129-2*, (D) *NID2*, (E) *HOXA11*, (F) *CELF2*, (G) *CHFR* and (H) *FLOT1* relative promoter methylation levels in lung cancer (LCA), normal lung (NL), breast cancer (BRCA), normal breast (NBr), colorectal cancer (CRC), normal colorectal (CRN), prostate cancer (P) and normal prostate (CP) tissues. Mann-Whitney U Test between tumor and normal tissues, n.s. $p>0.05$, *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$. Red horizontal lines represent the median methylation level.

Since our main goal was to develop a test capable of simultaneously detecting LC, BrC, CRC and PCa, the biomarker performance of the "PanCancer" selected genes was assessed (Table 3). Remarkably, individually, all genes' promoter methylation levels were able to discriminate cancer with at least 43% sensitivity and 87% specificity. Also, gene

panels were further constructed to increase overall biomarker performance (Table 3). Indeed, when “PanCancer” panel was considered positive whenever one of the 4 genes was positive (allowing the maximal sensitivity), this panel was able to discriminate LC with 70% sensitivity and 86% specificity, BrC with 97% sensitivity and 87% specificity, CRC with 93% sensitivity and 100% specificity and PCa with 70% sensitivity and 86% specificity. Besides, 100% specificity was also achieved for LC, BrC and PCa when two positive genes were considered for panel positivity (corresponding to 60%, 90%, and 60% sensitivity, respectively).

Table 3. Biomarker performance of each gene promoter methylation and “PanCancer” panel for lung, breast, colorectal and prostate cancer detection in tissue samples.

	Lung Cancer			Breast Cancer			Colorectal Cancer			Prostate Cancer		
	Cut-off	**SE%**	**SP%**	**Cut-off**	**SE%**	**SP%**	**Cut-off**	**SE%**	**SP%**	**Cut-off**	**SE%**	**SP%**
ADCY4	0,2155	57	93	61,2954	84	**100**	7,5524	66	**100**	27,267	93	**100**
MIR129-2	0,9647	50	93	0,6964	90	93	7,4221	86	**100**	1,9257	90	**100**
NID2	2,597	53	**100**	3,1411	68	**100**	4,3836	79	**100**	45,0803	97	87
MAGI2	7,2989	43	**100**	8,889	52	93	48,5787	45	**100**	15,6849	50	100
PANCANCER	1 POS. GENE	70	86	1 POS. GENE	97	87	**1 POS. GENE**	**93**	**100**	1 POS. GENE	70	86
	2 POS. GENES	**60**	**100**	**2 POS. GENES**	**90**	**100**	2 POS. GENES	86	100	**2 POS. GENES**	**60**	**100**
	3 POS. GENES	43	100	3 POS. GENES	68	100	3 POS. GENES	69	100	3 POS. GENES	43	100
	4 POS. GENES	30	100	4 POS. GENES	39	100	4 POS. GENES	47	100	4 POS. GENES	30	100

Abbreviations: SE% - Sensitivity; SP% - Specificity; 1 POS. GENE – at least one positive gene; 2 POS. GENES – at least two positive genes; 3 POS. GENES – at least three positive genes; 4 POS. GENES – at least four positive genes

III.AIM

BrC, LC, CRC and PCa are the most incident and among the deadliest cancers worldwide. Considering the limitations of currently available screening methods, there is a great need for the development of effective, accurate and minimally invasive screening procedures for cancer early detection. Indeed, DNA promoter hypermethylation appears to be a potential screening marker once it emerges early in carcinogenesis, its cancer- and tissue-specific, it is easy to assess in body fluids and it allows the detection of several cancer types in one single sample.

Thereby, the aim of this dissertation is to develop a minimally-invasive methylation-based test for simultaneous detection of BrC, LC, CRC and PCa.

Specifically, we intend to:

1. Assess the promoter methylation levels of new “CancerType” genes in tissue samples by multiplex qMSP, aiming to improve the discrimination capacity of the first tested genes (shown in preliminary data).
2. Evaluate different methodologies for cfDNA extraction from plasma samples and select the most suitable one.
3. Optimize methylation-specific droplet digital PCR (ddPCR) assays for evaluating promoter methylation levels of *ADCY4, MAGI2, MIR129-2* and *NID2*.
4. Assess the promoter methylation levels of *ADCY4, MAGI2, MIR129-2* and *NID2* by ddPCR in a set of plasma samples from patients with BrC, LC, CRC, PCa and asymptomatic controls (AC).
5. Evaluate the diagnostic performance of gene promoter methylation levels for each cancer type.
6. Evaluate the association between promoter methylation levels of each gene and clinicopathological features of each cancer type.

IV.MATERIAL AND METHODS

1. Clinical Samples

1.1 Tissue samples

Tissue samples from 120 cancer patients (31 BrCa; 30 PCa; 29 CRC; 30 LCa) submitted to surgery at IPO Porto, were selected. Additionally, as control samples, were selected: 16 (NL) lung tissue samples obtained from patients who underwent surgery due to benign conditions; 15 (NBr) breast tissue samples obtained from reduction mammoplasty of the contralateral breast of BrC patients without BrC hereditary syndrome; 15 (CRN) colorectal tissue samples obtained from surgical margins of non-gastrointestinal tumors with a total absence of tumor tissue; 15 (CP) prostate tissue samples obtained from patients who underwent radical cystoprostatectomy. All lung and colorectal samples corresponded to formalin-fixed paraffin-embedded (FFPE) tissues archived at the Department of Pathology of IPO Porto, while breast and prostate samples were fresh-frozen tissues, immediately frozen at -80ºC at the institutional tumor bank after surgical resection and examination (CES-IPOFG-EPE 120/015).

1.2 Plasma samples

For the optimization phase of this study, a total of 20 pre-treatment blood samples were selected, 4 for each tumor model (one of each stage) and 4 samples from asymptomatic blood donors (AC) (2 males and 2 females).

For the testing set, pre-treatment blood samples were collected from 200 patients diagnosed with LC (n=50), BrC (n=50), CRC (n=50) or PCa (n=50) between 2017 and 2020 at IPO Porto, Portugal. Additionally, for control purposes, blood samples were donated by 50 AC, from 2020 to 2022, at the same institution.

After collection of peripheral blood into EDTA-containing tubes, plasma was separated by centrifuging at 2500g for 30 minutes at 4ºC, and subsequently, stored at -80ºC in the institutional tumor bank until further use. All blood samples were processed within a maximum of 4h from the collection. Relevant clinical and pathological data were retrieved from clinical charts and an anonymized database was constructed for analysis purposes.

This study was approved by the institutional review board (Comissão de Ética para a Saúde) of IPO Porto, Portugal (CES-IPOFG-EPE 120/015). Written informed consent, following the Declaration of Helsinki ethical principles, was provided by all patients.

2. DNA extraction from tissue samples

For FFPE tissue samples, representative areas were macrodissected from 8 µm thick tissue sections to maximize the proportion of malignant cells (>70%), and subsequently deparaffinized, rehydrated using xylene and graded ethanol, and digested with proteinase K. For fresh-frozen tissue samples, 10 µm thick sections were cut and digested with SE buffer, 10% SDS and proteinase K (NZYTech, Portugal). Then, DNA was extracted from all samples using standard phenol-chloroform, ethanol, ammonium chloride and glycogen protocol. Briefly, subsequently to digestion, samples were transferred to Phase Lock Light 2mL tubes and mixed with 500 µL of phenol-chloroform pH=8 (PC8; Sigma-Aldrich, USA), followed by centrifugation at 13000 rpm for 15 minutes at 4ºC. Then, the aqueous phase was transferred to a new 2 mL tube, and the process was repeated. Afterward, the aqueous phase was transferred to a Safe Lock 2 mL tube and DNA was precipitated by mixing with cold absolute ethanol (2x the volume of DNA) (Merck Millipore, Germany), ammonia acetate at 7.5M (1/3 the volume of DNA) (Sigma-Aldrich, USA) and 2 µL of glycogen and left incubating overnight at -20ºC. Tubes were then centrifuged at 13000 rpm for 20 minutes and the supernatant was discarded. 1 mL of 70% cold ethanol was added and tubes were again centrifuged at 13000 rpm for 20 minutes. This washing step was repeated twice.

In the end, air-dried pellets were eluted in sterile bi-distilled water. All DNA elutes were stored at -20ºC until further use.

3. cfDNA extraction from plasma samples

cfDNA extraction from plasma samples was compared across 3 different methodologies: Manual extraction using QIAmp MinElute ccfDNA kit (Qiagen, Germany); manual extraction using MagMAX™ Cell-Free DNA Isolation Kit (Applied Biosystems, USA); automatic extraction using MagDEA® Dx SV kit for the magLEAD® 12gC extractor (Precision System Science, Japan). All protocols were performed according to manufacturer's instructions.

For the testing set samples, cfDNA was extracted from 2 – 4 mL of plasma using MagMAX™ Cell-Free DNA Isolation Kit. Plasma samples were prior centrifuged at 16,000g for 10 minutes at 4ºC, to remove any cellular debris. Then, the plasma supernatant was transferred to a 15 mL tube and MagMAX™ Cell Free DNA Magnetic Beads and MagMAX™ Cell Free DNA Lysis/Binding Solution were added in the appropriate ratio (Table 4). After a 10 minutes incubation period with shaking at 300rpm to bind cfDNA to the magnetic beads, the tubes were placed in a magnetic rack until the solution was clear and the beads were pelleted against the magnet. The supernatant was discarded, and 1 mL of MagMAX™ Cell Free DNA Wash Solution was added to wash the beads, with the bead slurry being then

transferred to a 1.5 mL tube and placed on a 1.5 mL tube-magnetic rack. After the solution was clear and the beads were pelleted against the magnet, the supernatant was discarded, and the washing step was repeated. Afterwards, washing with 80% ethanol was performed twice, and the beads were left to air dry for 5 minutes while still on the magnet. In the end, 22 µL of MagMAX™ Cell Free DNA Elution Solution were added to the beads, to elute the cfDNA, vortexed for 5 minutes and placed on the magnetic stand, with the supernatant containing the purified cfDNA.

All steps were performed at room temperature. The extracted cfDNA was stored at -20ºC until further use.

Table 4. Binding Solution/Beads Mix components.

Plasma (mL)	MagMAX™ Cell Free DNA Lysis/Binding Solution (mL)	MagMAX™ Cell Free DNA Magnetic Beads (µL)
2	2	20
2.5	2.5	25
3	3	30
3.5	3.5	35
4	4	40

4. DNA quantification

DNA extracted from tissue samples was quantified using NanoDrop Lite Spectrophotometer (Nanodrop Technologies, USA). 1 µL of DNA from each sample was used for such purpose.

DNA extracted from plasma samples was quantified in Qubit 4 Fluorometer using Qubit 1X dsDNA HS Assay Kit (Invitrogen, USA). 2 µL of each cfDNA sample and 198 µL of Qubit working solution were mixed in a 0.5 mL tube and used for quantification.

5. cfDNA fragmentation analysis

cfDNA fragment sizing and quantification was performed using the Cell-free DNA ScreenTape assay for 4200 TapeStation instrument (Agilent Technologies, USA). Such system performs electrophoretic separation of input cfDNA, allowing the identification of cfDNA subcomponents, their size and quantification, as well as total DNA concentration and the presence of contaminant high molecular weight DNA. Also, the assay provides a %cfDNA quality metric for evaluating the quality of cfDNA for downstream applications [132].

A ladder was prepared by mixing 15 µL of Cell-free DNA Sample Buffer and 15 µL of Cell-free DNA Ladder in a 0.5 mL tube (volume required for analyzing more than 2 ScreenTape devices). Also, 2 µL of cfDNA from each sample and 2 µL of Cell-free DNA Sample Buffer were pippeted into a 96-well plate (Agilent Technologies, USA), followed by vortexing at 2000 rpm for 1 min and spinning down. Results were analyzed with the TapeStation Analysis software version 4.1.1 (Agilent Technologies, USA).

6. Sodium-bisulfite modification

Sodium-bisulfite modification is a gold standard procedure regarding DNA methylation studies. This technique allows the differentiation of methylated from unmethylated cytosines by converting unmethylated cytosines into uracils, while methylated cytosines remain 5-methylcytosines. The conversion occurs by a series of chemical reactions: (1) Sulphonation: addition of bisulfite to the 5-6 double bond of cytosine; (2) Deamination: hydrolytic deamination of the resulting cytosine-bisulfite derivative to give an uracil-bisulfite derivative; (3) Desulphonation: removal of the sulphonate group to originate an uracil residue [133].

All DNA samples were modified using the EZ DNA Methylation-Gold™ Kit (Zymo Research, USA), according to the manufacturer's instructions. To begin, 130 µL of CT conversion reagent solution was added to the volume equivalent to 1000 ng of the previously extracted/quantified DNA samples. Regarding cfDNA, a total of 20 µL were used for the modification protocol. Then, samples were incubated at 98ºC for 10 minutes (DNA denaturation), followed by 64°C for 180 minutes (bisulfite conversion) in the Veriti 96-Well Thermal Cycler (Thermo Fisher Scientific, USA). Next, each sample was added to Zymo-Spin™ IC columns, plus 600 µL of M-Binding Buffer, and incubated for 10 minutes. The columns were centrifuged at 10,000 rpm for 30 seconds. After, 100 µL of M-Wash Buffer were added, followed by centrifugation. Next, 200 µL of M-Desulphonation Buffer were added to the column, followed by 20 minutes incubation and centrifugation. Following, the column was washed twice with 200 µL of M-Wash Buffer and centrifuged. Finally, the columns were transferred to 1.5 mL safe-lock tubes and sterile bi-distilled water was added to elute the modified DNA. After 5 minutes of incubation, the columns were centrifuged at 12,000 rpm for 30 seconds. This process was repeated twice. In the end, 60 µL of bisulfite-converted tissue-extracted DNA and 20 µL of bisulfite-converted cfDNA were obtained. Additionally, 1 µg of Human Methylated & Non-methylated DNA (Zymo Research, USA) were also modified using the previously mentioned protocol and eluted in 30 µL of sterile bi-distilled water.

All bisulfite-converted DNA was stored at -80ºC until further use.

7. Target-specific preamplification

DNA preamplification is performed when the available input DNA for downstream applications is limited. For such purpose, the SsoAdvanced™ PreAmp Supermix (Bio-Rad, USA) was used.

A preamplification assay pool was prepared by mixing 2.5 µL of forward and reverse primers at 100 µM of the interest genes and sterile bi-distilled water up to a volume of 500µL. For the reaction mix, 8 µL of sodium-bisulfite modified cfDNA were mixed with 25µL of SsoAdvanced PreAmp Supermix, 5 µL of the described assay pool and 12 µL of sterile bi-distilled water. Then, samples were incubated at 95ºC for 3 minutes, followed by 12 cycles of 95ºC for 15 seconds and 58ºC for 45 minutes in the Veriti 96-Well Thermal Cycler (Thermo Fisher Scientific, USA).

The resulting amplified cfDNA samples were diluted in sterile bi-distilled water in a 1:5 ratio prior to ddPCR.

8. Methylation-specific primer and probe design

Genomic DNA sequences of the target genes (plus 1500bp upstream – promoter region) were acquired from the UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly. The bisulfite-treated methylated DNA sequence was obtained through Methyl Primer Express v1.0. Specific forward/reverse primers and probes were designed to accommodate the CpG sites relevant for the study (as determined by the previous *in silico* analysis). FAM, HEX and Cy5 (for β-Actin) fluorochromes and BHQ quenchers were selected for each probe to allow the assessment of multiple genes simultaneously (multiplex PCR). Importantly, to assure optimal primer/probe properties, sequences were analyzed using the Primer Express 3.0 – Primer Probe Test Tool, and additionally with the Beacon Designer, Premier Biosoft. Finally, to assure specificity for only one (specific) PCR product, primer sequences were run through Bisearch Primer Design and Search Tool.

9. Quantitative methylation-specific PCR (qMSP)

Promoter methylation levels of *EDNRB, ZSCAN1, GFRA1* and *AOX1* were evaluated by multiplex qMSP, allowing the assessment of multiple genes simultaneously. β-Actin was used as the internal reference gene to normalize the assay.

Primers and probe sequences for each gene are listed in Table 5. The multiplex gene panels used are displayed in Table 6.

qMSP assays were run in 384-well plates using a QuantStudio™ 12K Flex Real-Time PCR System (Applied Biosystems, USA). Per each well, it was added 1 of µL bisulfite-modified DNA extracted from tissue samples, 5 µL of Xpert Fast Probe MasterMix (GRiSP, Porto, Portugal) with ROX, a variable volume of primers and probe at 10 µM (Table 5) and sterile bi-distilled water to a final volume of 10 µL.

The following PCR program was used: 1 cycle at 95ºC for 3 minutes (polymerase activation), 45 cycles at 95°C for 5 seconds (DNA denaturation) and a variable temperature, according to each panel (Table 6), for 30 seconds (annealing and extension).

All samples were run in triplicates. Three no template controls (NTC) and two negative controls (Human HCT116 DKO Non-Methylated DNA [Zymo Research, USA]) were included in every plate, assuring the absence of contaminations and specificity for the methylated DNA template. Serial dilutions (five, in duplicate) of a positive control (Human HCT116 DKO Methylated DNA [Zymo Research, USA]) were included in each plate, used to compute a standard curve and evaluate the efficiency of the run. All plates displayed efficiency values above 90%. Results were plotted as relative methylation levels (Target gene/β-Actin), multiplied by 1000 for easier tabulation.

Table 5. Primers and probes sequences with respective fluorochrome and quencher for qMSP.

Gene		Sequence (5'–3')	Volume	Vendor
β-Actin	Primers	F – TGGTGATGGAGGAGGTTTAGTAAGT R – ACCAATAAAACCTACTCCTCCCTTAA	0.8 µL	Sigma-Aldrich, Germany
	Probe	Cy5 – ACCACCACCCAACACACAATAACAAACACA – QSY	0.1 µL	Eurofins Genomics, Germany
EDNRB	Primers	F – GTCGTTTGGAGGGAATAGCGG R – CGAAAAACTCCTCCCGACG	0.8 µL	Eurofins Genomics, Germany
	Probe	HEX – TGCGGGTTTTCGAATTTTCGGCGTA – BHQ1	0.05 µL	Frilabo, Portugal
ZSCAN1	Primers	F – AATGTCGTCGTTTTGTTTCGC R – AATCGCCTTAACAACGAATCG	0.8 µL	Eurofins Genomics, Germany
	Probe	FAM – ATCGTATATGCGTATTTTCGTAGTCGT – BHQ1	0.05 µL	Frilabo, Portugal
GFRA1	Primers	F – GGCGGGAATAGGAGTAGGTCG R – CGAACAAAACCCTCGACTCG	0.8 µL	Eurofins Genomics, Germany
	Probe	FAM – TTCGGAATACGTTATTTTTCGCGTCGT – BHQ1	0.05 µL	Frilabo, Portugal
AOX1	Primers	F – GACGTTAAGCGTTATTGGCG R – TCTTCCCGAAACACCAACACG	0.8 µL	Eurofins Genomics, Germany
	Probe	HEX – CGTCGGATGATTTTCGTTTATATAGAGGGCGT – BHQ1	0.05 µL	Frilabo, Portugal

Table 6. Gene combinations for multiplex qMSP.

Combination 1	Annealing Temperature	Combination 2	Annealing Temperature
β-Actin		$GFRA1_{me}$	
$EDNRB_{me}$	60°C	$AOX1_{me}$	62°C
$ZSCAN1_{me}$		---	

10. Methylation-specific droplet digital PCR

Promoter methylation levels of *ADCY4, MIR129-2, NID2* and *MAGI2* were evaluated by multiplex droplet digital PCR (ddPCR). β-Actin was used as the internal reference gene to normalize the assay. Primers and probe sequences for each gene are listed in Table 7. The multiplex gene panels used are displayed in Table 8.

ddPCR reactions were prepared in 96-well plates. An assay mix (20X) was prepared for each gene panel by mixing the corresponding volume of primers and probes at 100μM (Table 7) and sterile bi-distilled water to the final volume, varying according to the number of samples being ran. Ten μL of sodium-bisulfite modified cfDNA, 1.1 μL of the prepared assay mix and 11 μL of 2X ddPCR Supermix for probe no dUTP (Bio-Rad, USA) were added per each well.

For droplet generation, 20 μL of the reaction mix and 70 μL of Droplet Generation Oil for Probes (Bio-Rad, USA) were pippeted into cartridges and then placed on the QX200 Droplet Generator (Bio-Rad, USA). After being generated, droplets were pippeted into a ddPCR 96-Well Plate (Bio-Rad, USA) and heat-sealed in a PX1 PCR Plate Sealer (Bio-Rad, USA) at 180ºC for 5 seconds. Subsequently, plates were placed on C1000 Touch Thermal Cycler (Bio-Rad, USA) and the following PCR program was used: 95ºC for 10 minutes, 50 cycles of 94ºC for 30 seconds and 57ºC (optimized annealing temperature) for 1 minute with a ramp rate of 2ºC/s, and 98ºC for 10 minutes. In the end, droplets where then read on the QX200 Droplet Reader (Bio-Rad, USA) to count droplets containing amplified target DNA and empty ones based on fluorescence. Positive, negative and no template controls were included in every plate. Additionally, to minimize contaminations, PCR reaction preparation and droplet generation/reading were performed in separate dedicated rooms.

The limit of blank (LOB) and limit of detection (LOD) were calculated for each target using negative control replicates, according to [134, 135]. A target was considered positive when more than LOB droplets were detected. The limit of quantification (LOQ) was evaluated by serial diluting methylated with unmethylated control DNA in 5 ng of total DNA in each well, as reported by Yu et al. [136].

Results were analyzed using the QuantaSoft Analysis Pro software (Bio-Rad, USA). Positive droplets were manually identified based on positive control performance at each experiment. Wells reporting less than 10,000 read droplets were not considered for analysis. The number of target copies per 20 μL outputted by the software were normalized for the plasma volume of each sample. Normalized copy numbers of each target gene were divided by normalized β-Actin copies and multiplied by 100 to provide the methylation percentage of each target.

Table 7. Primers and probes sequences with respective fluorochrome and quencher for ddPCR.

Gene		Sequence (5'–3')	Concentration	Vendor
β-Actin	Primers	F – TGGTGATGGAGGAGGTTTAGTAAGT R – ACCAATAAAACCTACTCCTCCCTTAA	400µM	Sigma-Aldrich, Germany
	Probe	HEX – ACCACCACC – ZEN – CAACACACAATAACAAACACA – IBFQ	250µM	Integrated DNA Technologies, USA
ADCY4	Primers	F – AAAGGAGACGGGATTGTTAC R – AACCGAACGCCGAATTAC	400µM	Eurofins Genomics, Germany
	Probe	FAM – TTTAGGTGG – ZEN – GGTTCGTCGGGTC – IBFQ	250µM	Integrated DNA Technologies, USA
MIR129-2	Primers	F – GGAGTGGTGAGATTGAGTCG R – GACTTCTTCGATTCGCCG	400µM	Eurofins Genomics, Germany
	Probe	HEX – CGCGTTGGG – ZEN – GAGATTTAGTTTGTTC – IBFQ	250µM	Integrated DNA Technologies, USA
NID2	Primers	F – TCGTAATTTCGTTATTCGTTCGC R – CCCGCAAAATTTAAAACAACG	400µM	Eurofins Genomics, Germany
	Probe	HEX/FAM – CCGCAACGA – ZEN – CGAATACGACTACTAACCTACG – IBFQ	125µM	Integrated DNA Technologies, USA
MAGI2	Primers	F – GGATTTCGCGTTAGGACGTTC R – ACCTCTATACGACCGAACCGC	400µM	Eurofins Genomics, Germany
	Probe	FAM – CGCGCCTAA – ZEN – TACCACATCTCGAACTCTACG – IBFQ	250µM	Integrated DNA Technologies, USA

Table 8. Gene combinations for multiplex ddPCR.

Combination 1	Annealing Temperature	Combination 2	Annealing Temperature
β-Actin		$ADCY4_{me}$	
$MAGI2_{me}$	57˚C	$MIR129\text{-}2_{me}$	57˚C
---		$NID2_{me}$	

11. Statistical Analysis

Data was tabulated using Microsoft Excel and analyzed and plotted using GraphPad Prism 9 (GraphPad Software Inc., USA) and IBM SPSS Statistics version 26 (IBM-SPSS Inc., USA) for MacOS. Non-parametric tests were performed to compare methylation levels of each gene's promoter between cancer and control samples, in both tissue and plasma samples, and to evaluate associations with clinicopathological features. Mann-Whitney test was used for comparisons between two groups, while Kruskal-Wallis test was used for multiple groups, followed by Dunn's multiple comparison test for pairwise comparisons. Pearson correlation was performed for comparing the expected and observed number of methylated copies for determining LOQ, while Spearman correlation was applied to correlate the levels of methylation with patients and controls' ages. A result was considered statistically significant when p-value<0.05.

For each gene promoter, samples were classified as methylated or non-methylated based on the cut-off values set using Youden's J index (combining highest sensitivity and specificity) [137], through receiver operator characteristic (ROC) curve analysis. Validity estimates (sensitivity, specificity and accuracy) were determined to assess detection biomarker performance (Table 9). To improve detection performance of the selected genes, panels were constructed considering a positive result whenever at least one gene promoter was classified as methylated.

Table 9. Formulas for biomarker performance calculations.

Cancer *vs.* Control			**Validity Estimates**	
	Cancer	Control		
> Cut-off	A	B	**Sensitivity (%)**	(A/E) x 100
< Cut-off	C	D	**Specificity (%)**	(D/F) x 100
Total	E	F	**Accuracy (%)**	[(A+D) / (E+F)] x 100

V.RESULTS

1. Selection and validation of new CancerType genes in tissue sample set

Given the suboptimal results regarding the "CancerType" genes obtained in the preliminary data, we sought to select new genes that would discriminate between BrC, LC, CRC and PCa. For that, we have chosen the following genes in line for each cancer type depicted in the TCGA data mining (Table 2). Accordingly, we designed methylation-specific primers and probes for *EDNRB, ZNF529, GFRA1* and *HRASLS5*. After optimization, methylation specificity could not be obtained with the designed sequences for *ZNF529* and *HRASLS5*, hence, *ZSCAN1* and *AOX1* were further selected for discriminating LC and PCa, respectively.

EDNRB, ZSCAN1, GFRA1 and *AOX1* promoter methylation levels were compared between tumoral tissues of each cancer type and respective normal tissues (Figure 8). *EDNRB* and *ZSCAN1* promoter methylation levels were significantly higher in CRC tissue compared to normal tissue ($p<0.0001$) and in LC and BrC, in a lower extent. Given that the methylation status of these genes is supposed to be specific of BrC and LC, respectively, their value as "CancerType" genes is rather weak. *GFRA1* methylation levels were also significantly higher in CRC tissue ($p<0.0001$), its specific cancer type, nonetheless, significant differences were also observed between LC and normal tissues. Regarding *AOX1*, promoter methylation levels were significantly higher in PCa tissue compared to normal tissue ($p<0.0001$), as expected, but CRC tissues also displayed high methylation levels of this gene promoter.

Overall, these results, together with the preliminary ones, showed a limited capacity of the "CancerType" panel in discriminating between the four cancers. Hence, major improvements need to be done in order to develop a better algorithm to select genes for identifying a tumor's site of origin.

Importantly, given the promising results obtained for the "PanCancer" genes to simultaneously detect all 4 cancer types in study, we decided to proceed with liquid biopsy testing focusing only on this panel.

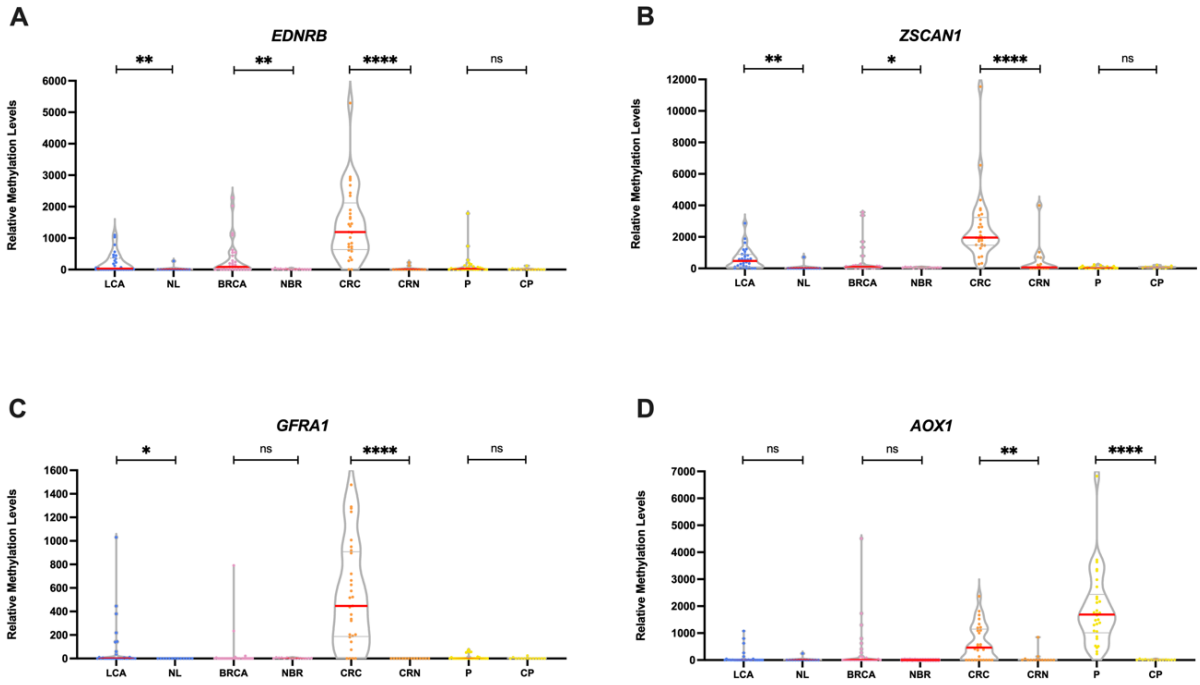

Figure 8. Distribution of (A) *EDNRB*, (B) *ZSCAN1*, (C) *GFRA1* and (D) *AOX1* relative promoter methylation levels in lung cancer (LCA), normal lung (NL), breast cancer (BRCA), normal breast (NBr), colorectal cancer (CRC), normal colorectal (CRN), prostate cancer (P) and normal prostate (CP) tissues. Mann-Whitney U Test between tumor and normal tissues, n.s. $p>0.05$, *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$. Red horizontal lines represent the median methylation level.

2. Clinical and Pathological Data

This study included 2 sets of plasma samples, an optimization and a testing set.

The optimization set consisted of 16 cancer patients (4 of each cancer type, one of each clinical stage) and 4 asymptomatic blood donors (AC), which served as controls.

Additionally, 200 cancer patients with BrC (n=50), LC (n=50), CRC (n=50), PCa (n=50) and 50 AC were selected for the testing set.

Detailed clinical and pathological characterization of the selected patients for both sets is provided in Table 10.

No correlation was observed between the methylation levels of any tested gene and the age of cancer patients or asymptomatic controls ($p>0.05$ for all genes).

Table 10. Clinical and pathological features of breast, lung, colorectal and prostate cancer patients and asymptomatic controls included in this study.

Clinicopathological features	
Optimization Set	
Breast Cancer	4 samples (Stage I-IV)
Lung Cancer	4 samples (Stage I-IV)
Colorectal Cancer	4 samples (Stage I-IV)
Prostate Cancer	4 samples (Stage I-IV)
Healthy Donors	4 samples (2 ♀ and 2 ♂)
Testing set	
Asymptomatic controls	
Number	50
Age median (range)	59 (40 – 66)
Breast Cancer	
Number	50
Age median (range)	59 (30 – 93)
Histological Type	
Lobular carcinoma in situ	1
Invasive lobular carcinoma	6
Ductal carcinoma in situ	1
Invasive ductal carcinoma	36
Other invasive carcinoma subtypes [a]	6
Estrogen Receptor Status	
Positive	47
Negative	3
Progesterone Receptor Status	
Positive	44
Negative	6
HER2 Expression Status	
Positive	6
Borderline	3
Negative	41
Primary Tumor (T)	
T1/T2	46
T3/T4	4
Regional Lymph Node (N)	
N0	33
N+	17
Distant Metastasis (M)	
M0	48
M+	2
Clinical Stage	
I/II	45
III/IV	5
Lung Cancer	
Number	50 (14 ♀, 36 ♂)
Age median (range)	63 (36 – 84)

Histological Type	
Non-small cell lung cancer (NSCLC):	
Adenocarcinoma	37
Squamous cell carcinoma	8
Large-cell neuroendocrine carcinoma	1
Small cell lung cancer (SCLC)	2
Carcinoid tumor	2
Primary Tumor (T) [b]	
T1	9
T2/T3/T4	27
Regional Lymph Node (N) [b]	
N0	17
N+	19
Distant Metastasis (M)	
M0	22
M+	28
Clinical Stage	
I/II	14
III/IV	36
Colorectal Cancer	
Number	50 (22 ♀, 28 ♂)
Age median (range)	62 (49 – 81)
Histological Type	
Adenocarcinoma	50
Tumor location	
Proximal colon	13
Distal colon	14
Rectum	23
Primary Tumor (T) [c]	
T1/T2	12
T3/T4	34
Regional Lymph Node (N) [c]	
N0	28
N+	18
Distant Metastasis (M) [c]	
M0	44
M+	2
Clinical Stage	
I/II	27
III/IV	23
Prostate Cancer	
Number	50
Age median (range)	66 (47 – 78)
Histological Type	
Adenocarcinoma	50
Primary Tumor (T) [d]	
T1/T2	26
T3/T4	23

Regional Lymph Node (N) [d]	
N0	38
N+	11
Distant Metastasis (M) [d]	
M0	42
M+	7
PSA Levels (ng/mL) [e]	
<10	32
10-20	9
>20	7
Clinical Stage	
I	8
II	19
III/IV	23

[a] includes NST (no special type), mucinous, pleomorphic and mixed type carcinoma; [b] no information available in 14 cases; [c] no information available in 4 cases; [d] no information available in 1 case; [e] no information available in 2 cases.

3. Optimization of methylation-specific ddPCR assays

To optimize methylation-specific ddPCR assays, fully methylated and non-methylated DNA were used at a 1:100 dilution, to mimic cfDNA concentration. The optimization phase aimed:

1) At finding the adequate settings to allow the best separation between positive and negative droplet populations for each target and between droplet populations (in multiplex panels).
2) To accurately consider a sample as positive only in the presence of methylated DNA.

3.1 Optimal annealing temperature

A temperature gradient (54-62ºC) was performed to select the annealing temperature that provided the best separation between positive and negative droplets, as well as the highest amplitude for the positive population (Figure 9).

Except for *MAGI2*$_{me}$, in which the annealing temperature did not affect much the amplitude of positive droplets, an annealing temperature of 57ºC or below should be selected for the remaining genes. Moreover, as our goal was to further develop multiplex panels, we selected 57ºC as the annealing temperature for all genes, since less “droplet rain” between the 2 droplet clusters was attained, while providing the highest amplitude.

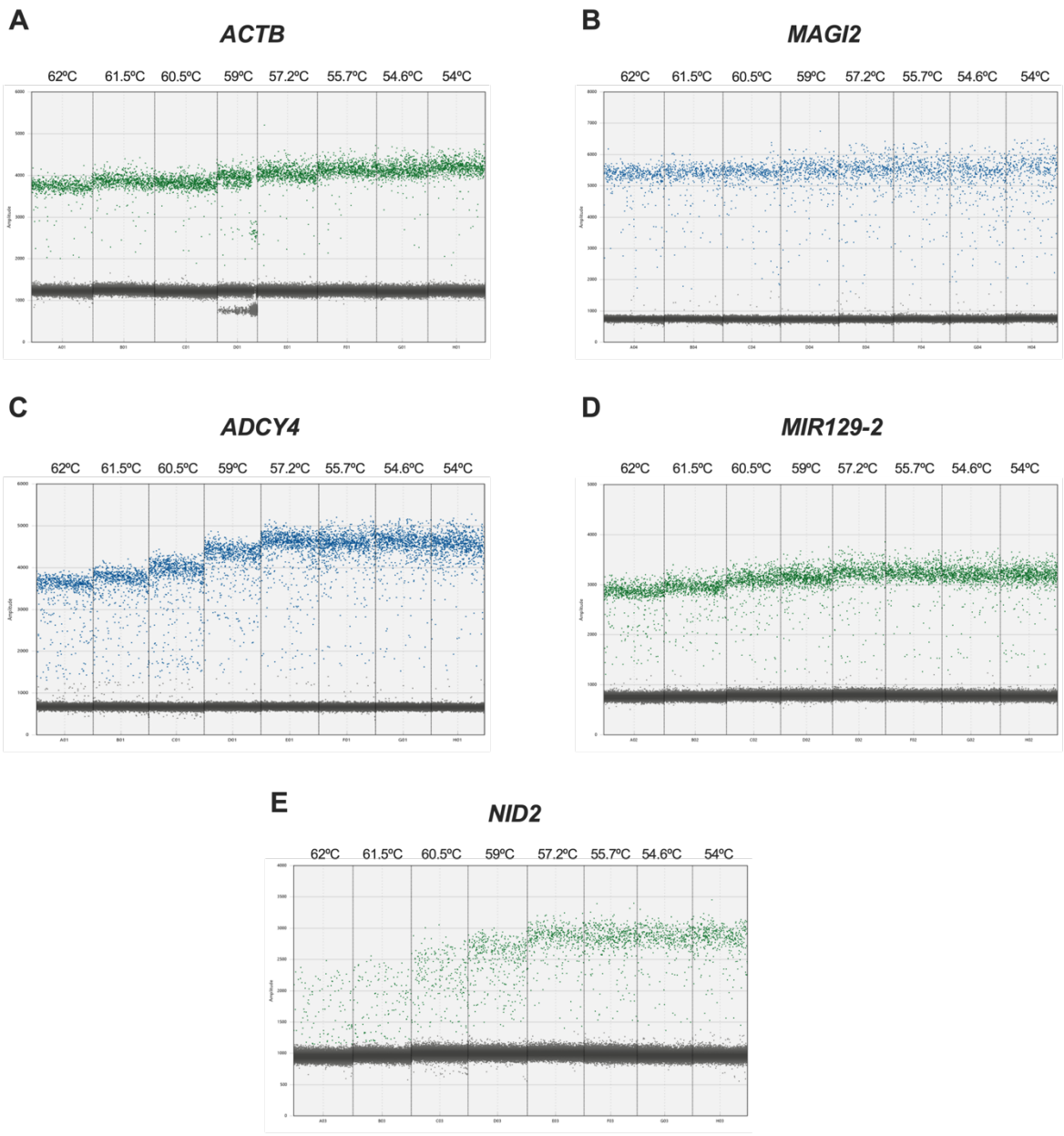

Figure 9. Temperature gradient for selection of the optimal annealing temperature for (A) *ACTB*, (B) *MAGI2*, (C) *ADCY4*, (D) *MIR129-2* and (E) *NID2* promoter's methylation. The highest separation between positive (green/blue) and negative (grey) droplets was achieved at 57ºC for all genes.

3.2 Multiplex panels

Given that the "PanCancer" panel consists of 5 genes (4 targets and reference gene), 2 individual panels (1 duplex panel and 1 triplex panel) were defined. Several primer and probe concentrations of the chosen genes and gene combinations were tested in order to select the best panels to be assessed simultaneously.

Considering that the available ddPCR system only has 2 fluorescence channels, the triplex panel was developed by using 2 probes for the same target labeled with different

fluorochromes at a 50:50 ratio to create a third droplet population with intermediate fluorescence amplitude between the other 2 targets (Figure 10B) [138].

The selected gene panels were a duplex combining *MAGI2*$_{me}$ and the reference gene *ACTB* (Figure 10A), and a triplex combining *ADCY4*$_{me}$, *MIR129-2*$_{me}$ and *NID2*$_{me}$ (Figure 10B). All genes were optimized with 400nM of each primer and 250nM of probe, except for *NID2*$_{me}$ with 125nM of both HEX/FAM-labeled probes.

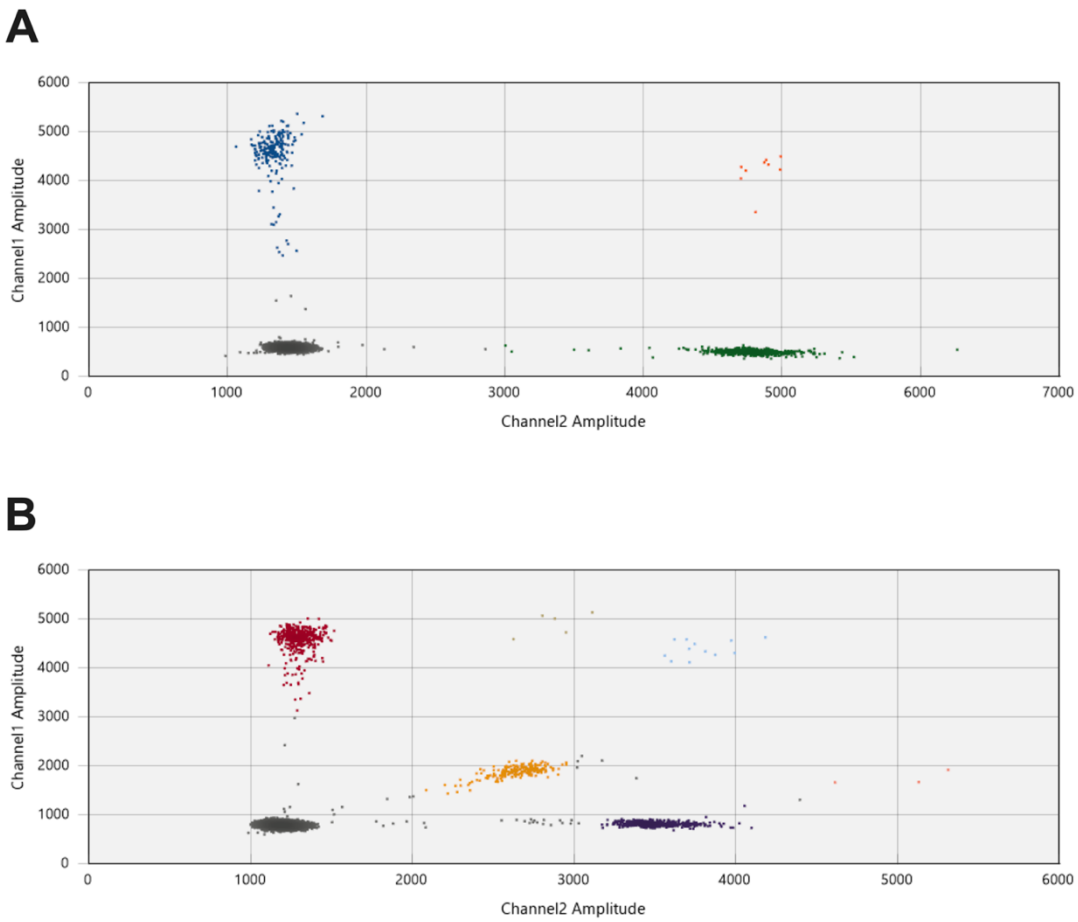

Figure 10. 2D-plots of the optimized multiplex gene panels. (A) Duplex panel combining *MAGI2*$_{me}$ (blue droplets) and *ACTB* (green droplets); (B) Triplex panel combining *ADCY4*$_{me}$ (red droplets), *MIR129-2*$_{me}$ (purple droplets) and *NID2*$_{me}$ (dark yellow droplets).

3.3 Limit of blank (LOB), detection (LOD) and quantification (LOQ)

For determining the LOB and LOD, consisting in the highest analyte concentration likely to be observed for a blank sample and the lowest amount of analyte that can be reliably detected, respectively, 30 replicates of negative (fully unmethylated DNA) control were ran for the 2 gene panels. The calculated LOBs (Table 11) of each gene were set as thresholds of positivity, meaning that a testing sample was considered positive for a certain gene if it showed a number of positive droplets higher than the respective LOB.

Table 11. Limit of blank (LOB) and limit of detection (LOD) for *ADCY4, MIR129-2, NID2* and *MAGI2*. Values are displayed as number of positive droplets.

Methylated Gene Promoter	LOB	LOD
ADCY4	0	3
MIR129-2	0	3
NID2	3	7
MAGI2	0	3

For determining the LOQ, defined as the lowest amount of analyte that can be reliably quantified, positive control (fully methylated DNA) was serially diluted with negative control (fully unmethylated DNA) from 100% to 0% and used as the template for each gene panel run (Figure 11). The expected number of methylated copies per gene was correlated with the observed number of copies. For all genes, a near perfect correlation was observed, with Pearson correlation coefficients above 0.994 for *ADCY4, MIR129-2* and *MAGI2*, and 0.9893 for *NID2* ($p<0.0001$ for all genes). Thereby, the assays can accurately quantify methylation even with 90% of non-methylated DNA background.

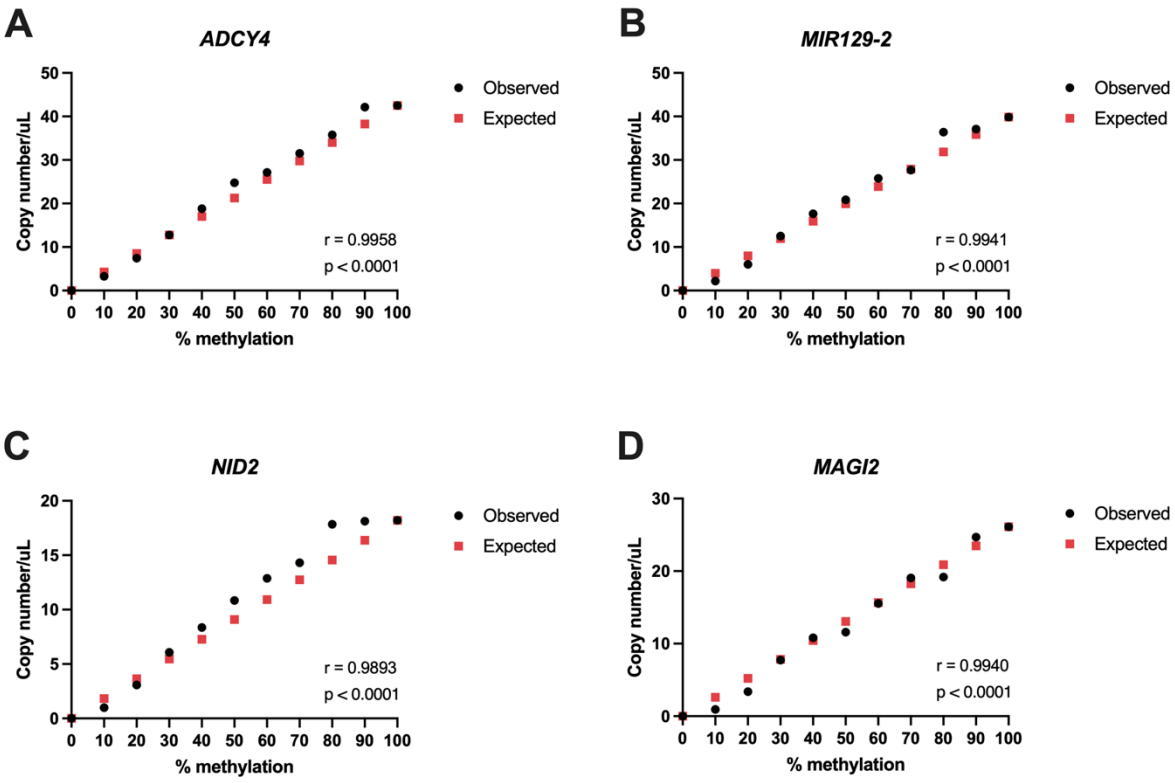

Figure 11. Limit of quantification (LOQ) for (A) $ADCY4_{me}$, (B) $MIR129\text{-}2_{me}$, (C) $NID2_{me}$ and (D) $MAGI2_{me}$. Observed number of methylated copies were correlated with the expected copy number for different methylation percentages. r – Pearson correlation coefficient.

4. Optimization of a pipeline for cfDNA downstream methylation analysis

4.1 Selection of an optimal method for cfDNA extraction from plasma samples

A set of 20 plasma samples (optimization set; Table 10) was selected to compare the performance of 3 different methods for cfDNA extraction. Comparison was performed by evaluating cfDNA profiles of each sample extracted by each methodology in the TapeStation system.

The average size of cfDNA is around 170bp corresponding to mononucleosomal DNA, but longer fragments can also be identified, consisting of di- and tri-nucleosomes, however in less abundance [132]. Thereby, in an electrophoretic separation of cfDNA it is expected to observe a prominent peak at around 170bp and smaller peaks in the area between 300-500bp. DNA fragments with size above 700bp are considered high molecular weight (HMW) DNA, most likely contaminant DNA arising from white blood cells [139].

Representative electropherograms from a stage IV LC sample extracted with the 3 kits are provided in Figure 12.

Samples extracted with both QIAmp MinElute ccfDNA Kit and MagMAX™ Cell-Free DNA Isolation Kit displayed expected cfDNA fragmentation profiles, all presenting a peak in the region between 160-180bp. Additionally, all samples presented one additional lower intensity peak around 400bp, while most even displayed a third peak near 500bp, showing that these extraction kits can maintain the integrity and extract all cfDNA subpopulations (Figure 10A, B). Concerning samples extracted with the magLEAD® 12gC automatic extractor, no prominent peaks were observed within 50-700bp, suggesting that such methodology results in a highly fragmented sample, thus, not being suitable for downstream molecular analysis (Figure 12C). Besides, it was possible to observe that when HMW DNA was present, its peak intensity was higher in samples extracted with QIAmp MinElute ccfDNA Kit (Figure 12A, B), indicating that this kit may yield a more contaminated sample.

Beyond fragment size analysis, the TapeStation system also allowed cfDNA populations' quantification (Figure 13). While no differences were depicted in total DNA nor DNA sized between 50-700bp (range considered cfDNA by the analysis software) concentrations obtained with QIAmp MinElute ccfDNA Kit and MagMAX™ Cell-Free DNA Isolation Kit, significantly lower concentrations were obtained when using magLEAD® 12gC extractor ($p<0.0001$) (Figure 13A, B). Regarding %cfDNA, a metric automatically computed by the software as a sample quality check, significantly lower values were also displayed for samples extracted automatically ($p<0.0001$) (Figure 13C).

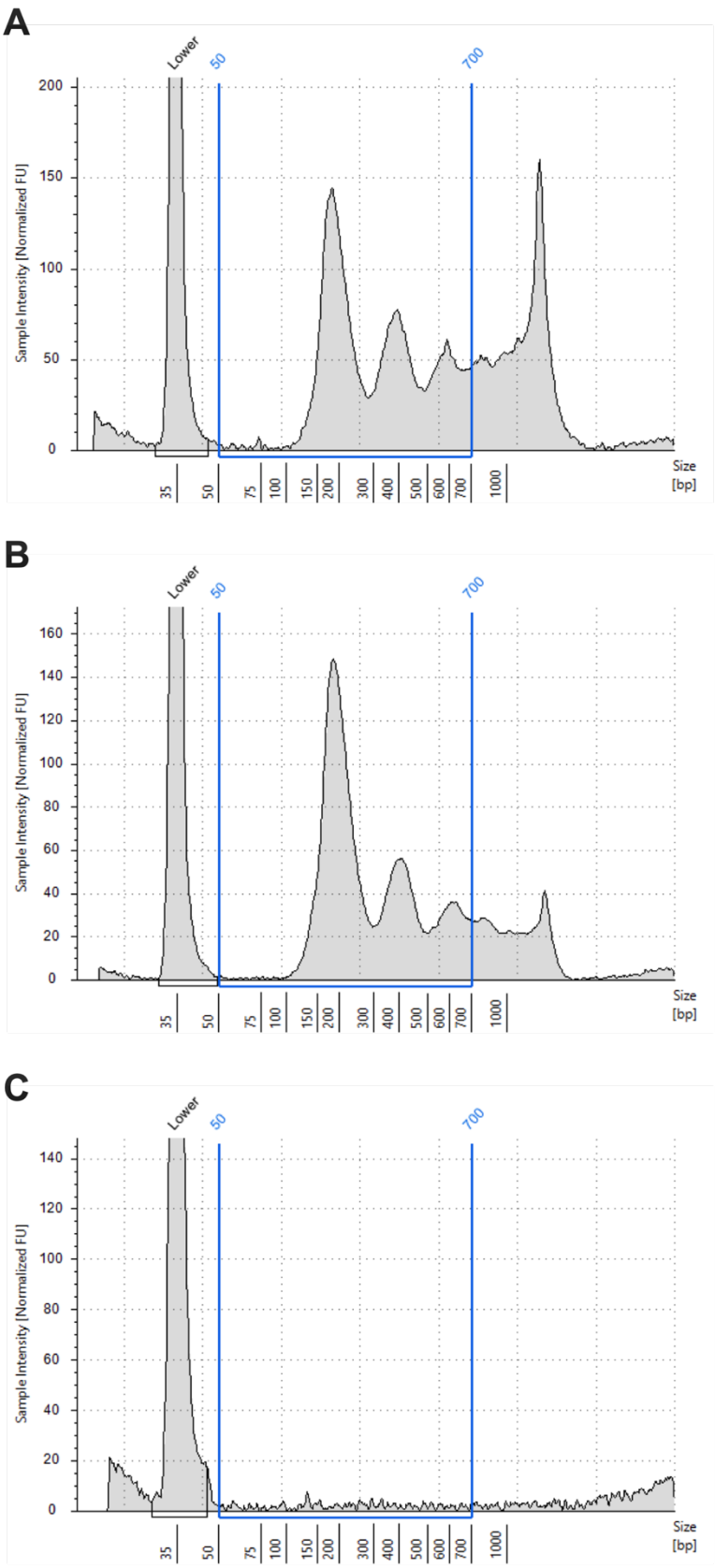

Figure 12. Electropherogram profiles of cfDNA extracted from a stage IV lung cancer plasma sample with (A) QIAmp MinElute ccfDNA kit, (B) MagMAX™ Cell-Free DNA Isolation Kit and (C) magLEAD® 12gC extractor. Obtained from TapeStation Analysis software.

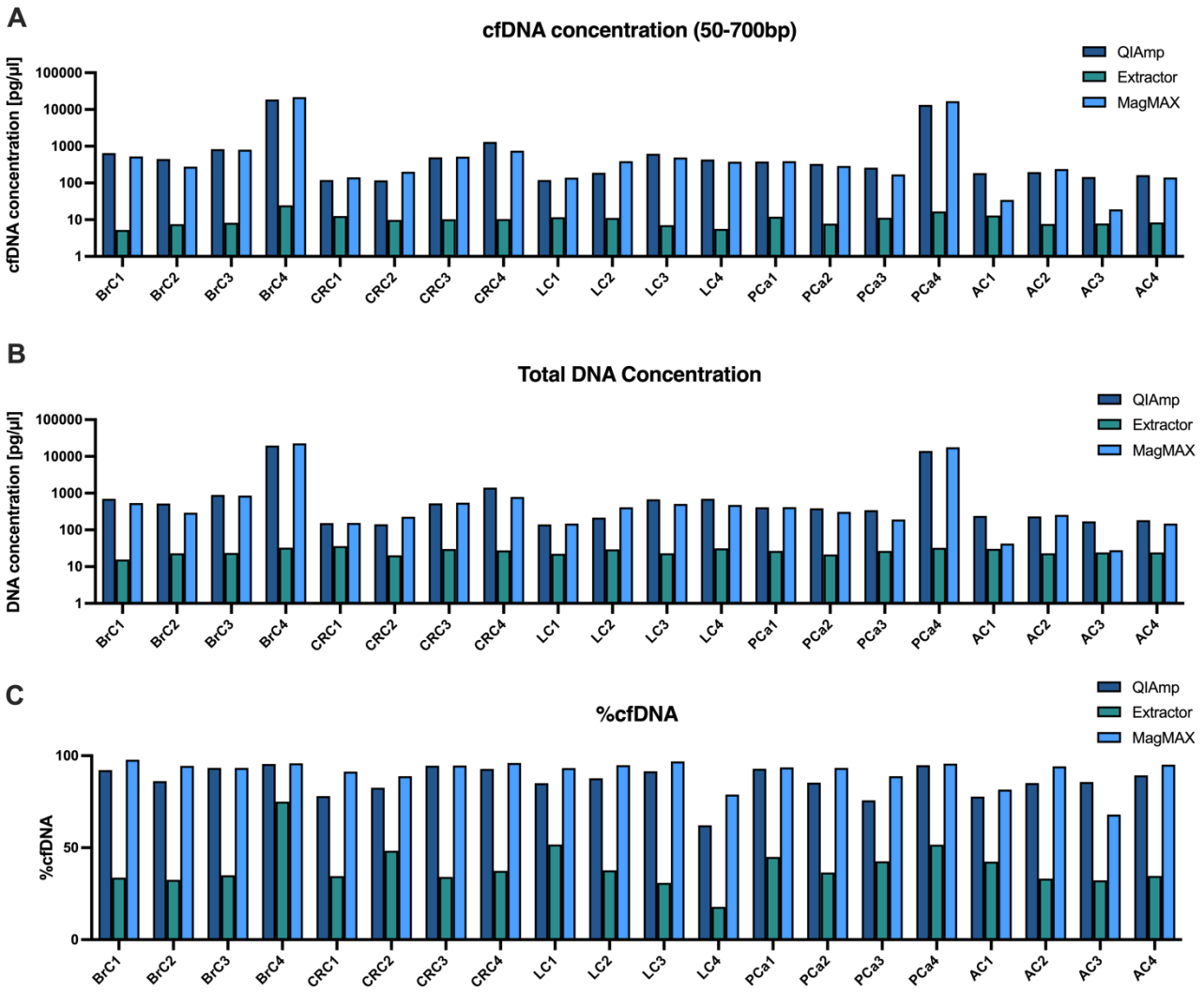

Figure 13. Comparison of cfDNA extraction using QIAmp MinElute ccfDNA kit, MagMAX^{TM} Cell-Free DNA Isolation Kit and magLEAD® 12gC extractor across 20 plasma samples. (A) Concentration of DNA with size ranging from 50-700bp. (B) Total DNA concentration. (C) %cfDNA i.e., % of DNA with 50-700bp in the sample. Numbers in sample name refer to cancer stage, for example, BrC1 corresponds to a stage I breast cancer sample. Data obtained from TapeStation Analysis software.

Although no significant differences were obtained in total or cfDNA concentrations between the samples extracted with the 2 manual kits (Figure 13A, B and Figure 14A), samples extracted with MagMAX™ Cell-Free DNA Isolation Kit showed significant higher %cfDNA (p=0.0069) (Figure 13C and 14B), indicating a higher proportion of contaminant HMW DNA present in samples extracted with QIAmp MinElute ccfDNA Kit. In fact, most samples extracted with MagMAX™ Cell-Free DNA Isolation Kit displayed %cfDNA above 90, emphasizing the quality of extraction.

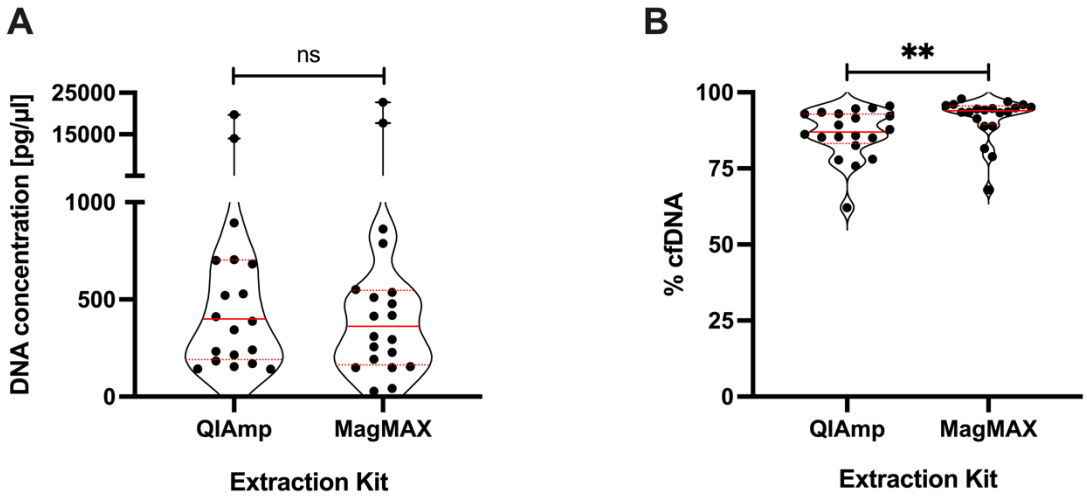

Figure 14. Comparison between cfDNA extraction using QIAmp MinElute ccfDNA kit and MagMAX™ Cell-Free DNA Isolation Kit. (A) Total DNA concentration of samples extracted with the 2 kits. (B) %cfDNA of samples extracted with the 2 kits. Mann-Whitney U Test between kits, n.s. $p>0.05$, **$p<0.01$. Red lines represent the median value and interquartile range. Data obtained from TapeStation Analysis software.

In addition to the quality of extracted samples, MagMAX™ Cell-Free DNA Isolation Kit also resulted in a higher cfDNA concentration when initial volumes were normalized, since this kit used 1mL of input plasma, while QIAmp MinElute ccfDNA Kit required a minimum of 2mL. Thus, MagMAX™ Cell-Free DNA Isolation Kit was selected for further experiments.

4.2 cfDNA input and cut-offs for sample eligibility

After fragmentation analysis, cfDNA samples were submitted to sodium-bisulfite modification and promoter methylation levels of the genes of interest were assessed by ddPCR, as previously described.

Surprisingly, only 3 samples showed positive droplets for any of the genes, being them stage IV BrC, CRC and PCa samples. Accordingly, these samples also showed a higher number of β-Actin copies/µL, indicating that more input DNA was present (Figure 15A).

To confirm that the obtained results were due to lack of input and not related to gene performance, we performed targeted pre-amplification on all samples to increase the

amount of DNA available for the ddPCR reaction. This time, all samples had positive droplets for at least one of the genes, except stage II LC and PCa samples (Figure 15B).

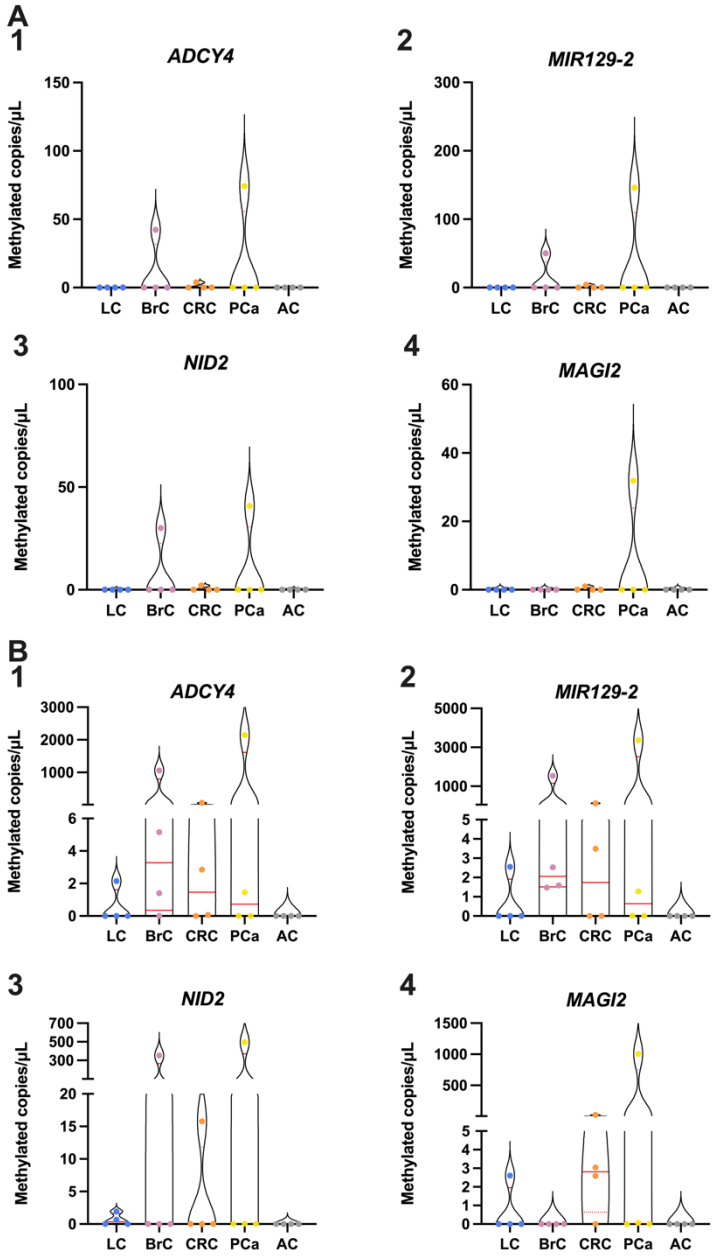

Figure 15. Number of methylated copies/µL for *ADCY4*, *MIR129-2*, *NID2* and *MAGI2* promoters across 20 plasma samples using ddPCR (A) and targeted pre-amplification followed by ddPCR (B). Number of methylated copies/µL was computed automatically by the software after manual threshold setting. Red lines represent the median value and interquartile range.

Since insufficient input led to the low methylation levels depicted, we sought to look at the DNA quantity corresponding to the volume used in the ddPCR reaction (5 µL) for each sample (Figure 16). Indeed, all samples except the 3 showing methylated copies for any gene had an input below 5ng. Thus, we defined as a criterion for sample eligibility that only

samples with concentration sufficient for inputting 5ng of DNA in each gene panel would be considered for methylation analysis. Additionally, a cut-off of 250 copies/20µL was defined for β-Actin to minimize the rate of false negative results.

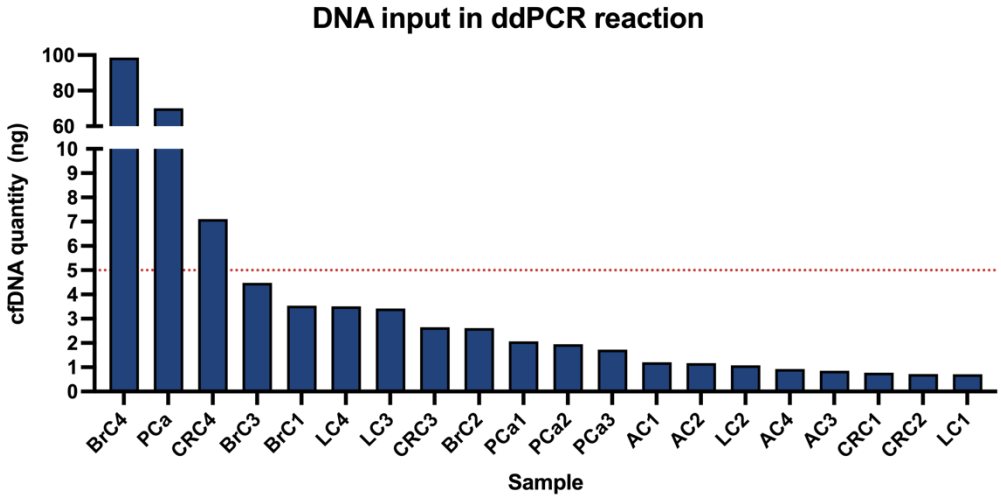

Figure 16. Input DNA in nanograms (ng) used in the ddPCR reaction for each sample. 5ng was defined as cut-off for sample quality control.

5. Gene promoter methylation levels in cfDNA

5.1 cfDNA concentrations across cancer patients and healthy donors

The median cfDNA concentration was higher in all four cancers compared to asymptomatic controls. Nonetheless, since different plasma volumes were used, concentration values were normalized to provide a more accurate comparison. After normalization, LC showed the highest median concentration, reaching up to 89.5ng of cfDNA per mL of plasma, followed by CRC, while BrC, PCa patients and AC displayed similar normalized median concentration values (Table 12 and Figure 17A).

Table 12. Concentration values of cfDNA extracted from plasma samples of lung, breast, colorectal and prostate cancer patients and asymptomatic controls included in this study.

Concentration	Lung Cancer	Breast Cancer	Colorectal Cancer	Prostate Cancer	Controls
Raw (ng/µL) Median (range)	0.911 (0.332-7.16)	0.752 (0.391-2.57)	1.325 (0.321-13.5)	0.749 (0.168-5.64)	0.684 (0.138-2.66)
Normalized (ng/mL) Median (range)	9.54 (2.36-89.5)	5.30 (3.12-14.25)	8.71 (2.35-84.86)	5.42 (1.23-35.45)	5.91 (1.2-16.72)

Overall, increased cfDNA concentration was observed across disease stages for all cancer types individually, except for BrC. Significant differences were observed between concentrations of early- and late-stage LC and CRC patients, while stage III PCa patients showed significantly higher cfDNA concentrations than stage I/II patients (Figure 17B).

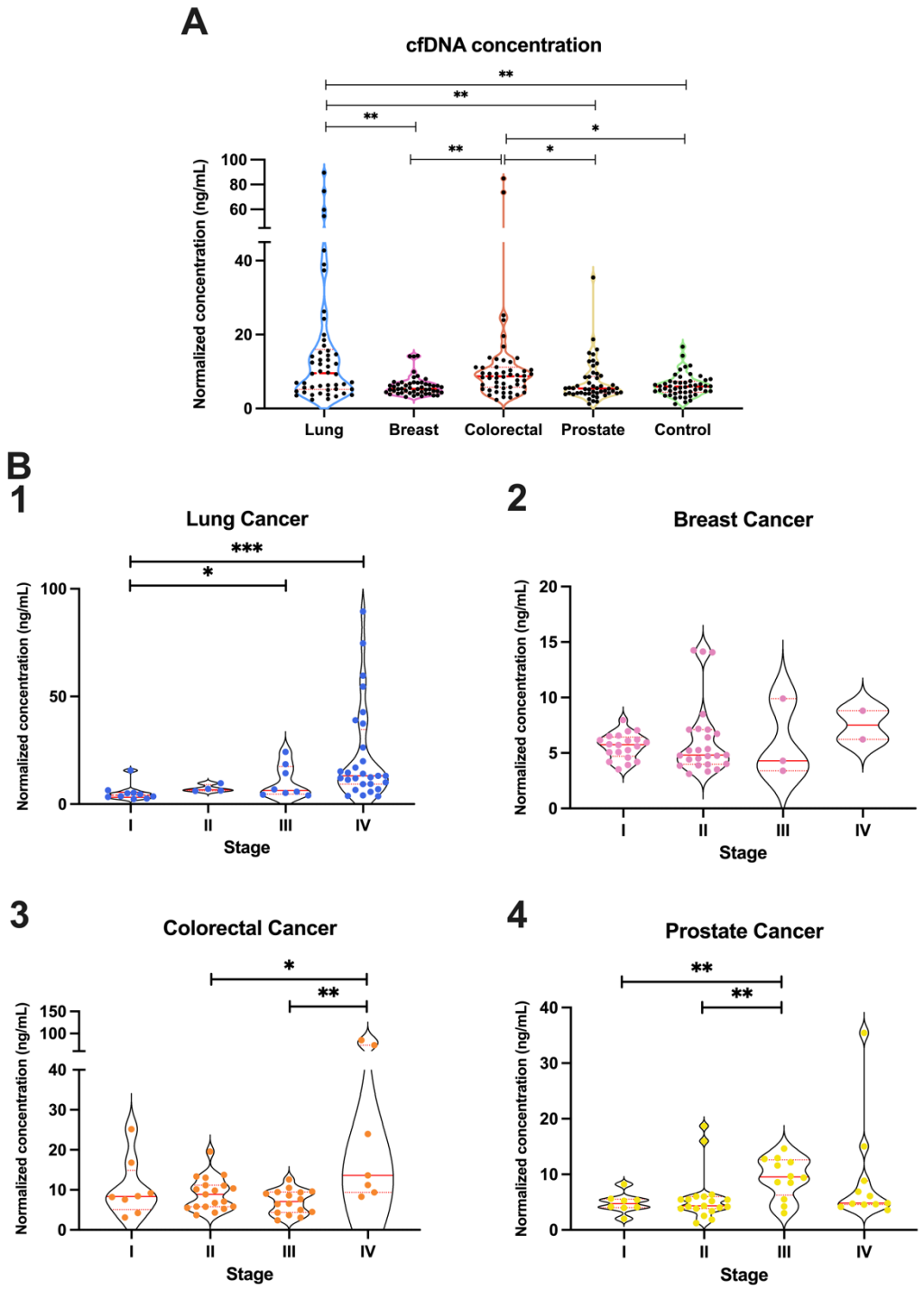

Figure 17. Normalized cfDNA concentration values per mL of plasma across lung, breast, colorectal and prostate cancer and asymptomatic controls (A) and between cancer stages (B). Kruskal-Wallis test followed by Dunn's multiple comparisons test between groups, *p<0.05, **p<0.01, ***p<0.001. Red lines represent the median value and interquartile range.

5.2 cfDNA methylation across cancer patients and healthy donors

cfDNA methylation levels of *ADCY4, MIR129-2, NID2* and *MAGI2* promoters were compared between each cancer type and controls (Figure 18). After applying the sample concentration and β-Actin copy number cut-offs for sample quality control, 38 LC samples, 25 BrC, 47 CRC, 39 PCa and 33 AC were considered suitable for methylation analysis.

All gene promoter methylation levels were significantly higher in LC patients compared to AC ($p<0.0001$ for $ADCY4_{me}$, $MIR129\text{-}2_{me}$ and $MAGI2_{me}$; $p<0.01$ for $NID2_{me}$). Concerning BrC, $ADCY4_{me}$ and $MAGI2_{me}$ levels were significantly higher ($p<0.01$), although no significant differences were apparent for $MIR129\text{-}2_{me}$ and $NID2_{me}$ ($p=0.1128$ and $p=0.1880$, respectively). Remarkably, significantly higher methylation levels were depicted for all gene promoters in CRC patients, with $p<0.0001$ for all genes. Similarly, all gene promoters showed higher methylation levels in PCa samples compared with AC, ($p=0.0167$ for $MIR129\text{-}2_{me}$, $p<0.01$ for $MAGI2_{me}$ and $ADCY4_{me}$ and $p<0.001$ for $NID2_{me}$).

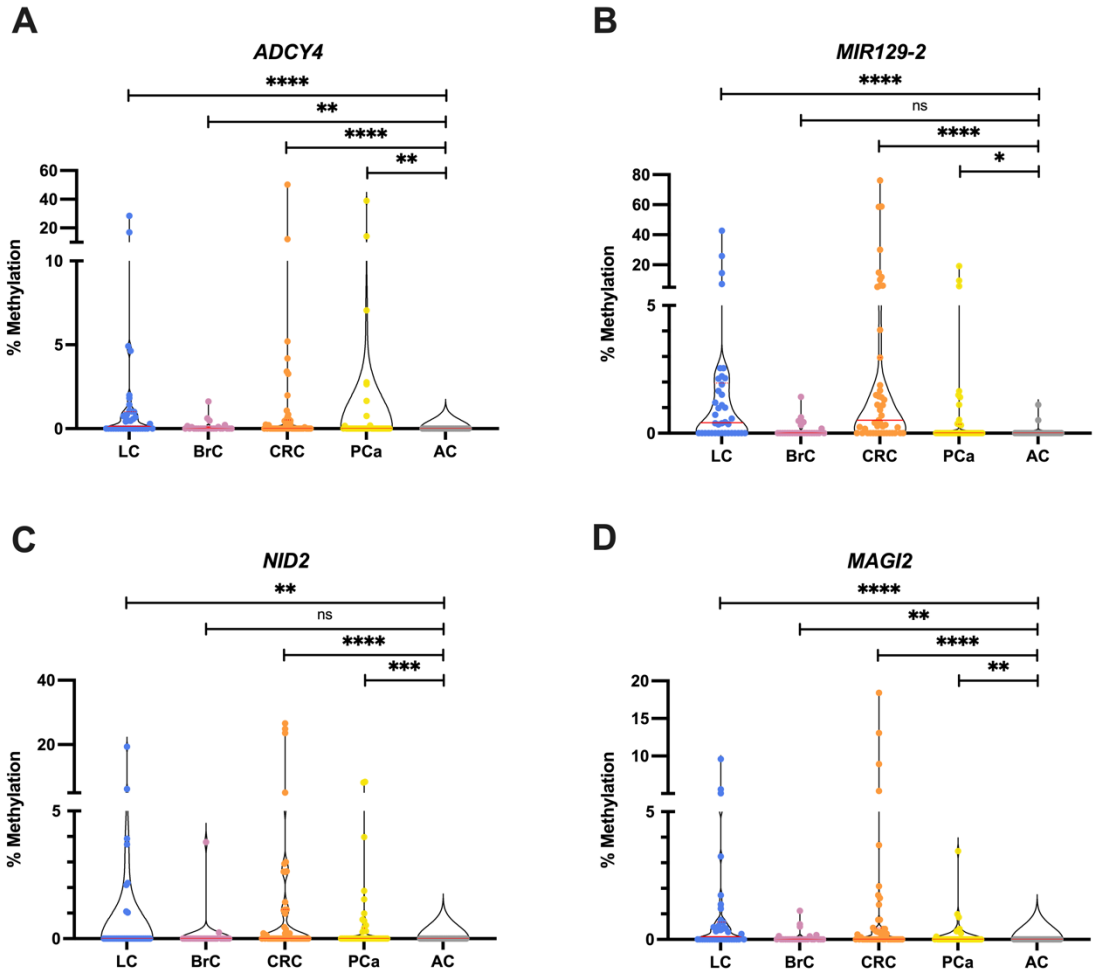

Figure 18. Distribution of (A) *ADCY4*, (B) *MIR129-2*, (C) *NID2* and (D) *MAGI2* promoter methylation levels in lung (LC), breast (BrC), colorectal (CRC) and prostate (PCa) cancers and asymptomatic controls (AC) samples. Mann-Whitney U Test between AC and each cancer type, n.s. $p>0.05$, *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$. Red horizontal lines represent median methylation levels.

5.3 Biomarker performance of gene promoter methylation levels

Since our main goal was to develop a test for simultaneous detection of BrC, LC, CRC and PCa, we further assessed the biomarker performance of the methylation levels of selected gene promoters in detecting individual cancer types, as well as pan-cancer detection.

Remarkably, 100% specificity was obtained for all genes, except for *MIR129-2*$_{me}$, in all cancer types (Table 13). In fact, no positive droplets were observed in AC for all the tested genes, except for *MIR129-2* in 2 samples. Nonetheless, a specificity of 93.75% was obtained for the methylation level of this gene promoter in discriminating each of the cancers from the controls.

Despite their high specificity, all genes displayed low sensitivity for BrC detection, with a maximum of 28% for *ADCY4*$_{me}$. Similarly, PCa detection reached a maximum sensitivity of 30.77% for *NID2*$_{me}$. While the methylation levels of this gene were only able to detect LC with 21.05% sensitivity, *ADCY4*$_{me}$ and *MAGI2*$_{me}$ showed 50% sensitivity and *MIR129-2*$_{me}$ 60.53%. Regarding CRC detection, all gene promoter methylation levels displayed sensitivities above 42%, reaching 68.09% for *MIR129-2*$_{me}$ levels (Table 13).

Gene panels were further constructed to increase detection sensitivity. Thus, the "PanCancer" panel combining all genes was applied and a sample was considered positive when at least one of the gene promoters presented methylation above the established cut-off. Using this approach, the 4 cancer types could be detected with 93.75% specificity, while sensitivities were 52% for BrC, 56.41% for PCa, 78.72% for CRC and 85.71% for LC detection (Table 14). 100% specificity was achievable by considering 2 or more positive genes, however, at the cost of reduced sensitivity (Table 14).

Given the importance of early cancer detection, we further assessed the performance of the "PanCancer" panel in detecting stage I/II cancer. Remarkably, early-stage cancer detection was possible with similar performance as for all stages, while stage III/IV detection sensitivity was slightly higher (Table 15).

When applied to the simultaneous detection of the 4 cancers, the "PanCancer" panel was able to correctly identify 104 out of 149 cancer patients, resulting in 69.80% sensitivity, 93.75% specificity and 74.03% accuracy (Table 16 and Figure 19).

Table 13. Biomarker performance of each gene promoter methylation levels for breast, lung, colorectal and prostate cancer detection.

Gene	Cut-off	Sensitivity%	Specificity%	Accuracy%
Breast Cancer				
ADCY4$_{me}$	0.045	28	100	68.42
MIR129-2$_{me}$	0.09	24	93.75	63.16
NID2$_{me}$	0.12	8	100	59.65
MAGI2$_{me}$	0.06	24	100	66.67
Lung Cancer				
ADCY4$_{me}$	0.145	50	100	72.86
MIR129-2$_{me}$	0.17	60.53	93.75	75.71
NID2$_{me}$	0.51	21.05	100	57.14
MAGI2$_{me}$	0.11	50	100	72.86
Colorectal Cancer				
ADCY4$_{me}$	0.04	42.55	100	65.82
MIR129-2$_{me}$	0.085	68.09	93.75	78.48
NID2$_{me}$	0.07	42.55	100	65.82
MAGI2$_{me}$	0.045	48.94	100	69.62
Prostate Cancer				
ADCY4$_{me}$	0.085	23.08	100	57.75
MIR129-2$_{me}$	0.155	28.95	93.75	58.57
NID2$_{me}$	0.13	30.77	100	61.97
MAGI2$_{me}$	0.055	20.51	100	56.34

Table 14. Biomarker performance of the "PanCancer" gene panel for breast, lung, colorectal and prostate cancer detection.

PanCancer	Positive genes*	Breast Cancer		Lung Cancer		Colorectal Cancer		Prostate Cancer	
		SE%	SP%	SE%	SP%	SE%	SP%	SE%	SP%
	1	52	93.75	85.71	93.75	78.72	93.75	56.41	93.75
	2	20	100	57.89	100	61.70	100	20.51	100
	3	5.88	100	34.21	100	42.55	100	17.95	100
	4	0	100	5.41	100	19.15	100	7.69	100

***** "PanCaner" panel is considered positive for a given sample if at least 1, 2, 3 or all 4 genes are positive.

Abbreviations: SE – sensitivity; SP – specificity.

Table 15. Biomarker performance of the "PanCancer" panel for early (stage I/II) and late (stage III/IV) stage detection of breast, lung, colorectal and prostate cancer.

Early-stage Cancer		
	Sensitivity%	**Specificity%**
Breast Cancer	52.17	93.75
Lung Cancer	77.78	93.75
Colorectal Cancer	76.92	93.75
Prostate Cancer	55	93.75
Late-stage Cancer		
	Sensitivity%	**Specificity%**
Breast Cancer	50	93.75
Lung Cancer	86.21	93.75
Colorectal Cancer	80.95	93.75
Prostate Cancer	57.89	93.75

Table 16. Biomarker performance of the "PanCancer" panel for simultaneous detection of breast, lung, colorectal and prostate cancer.

PanCancer – 4 Major Cancers	
Sensitivity	69.80%
Specificity	93.75%
Accuracy	74.03%

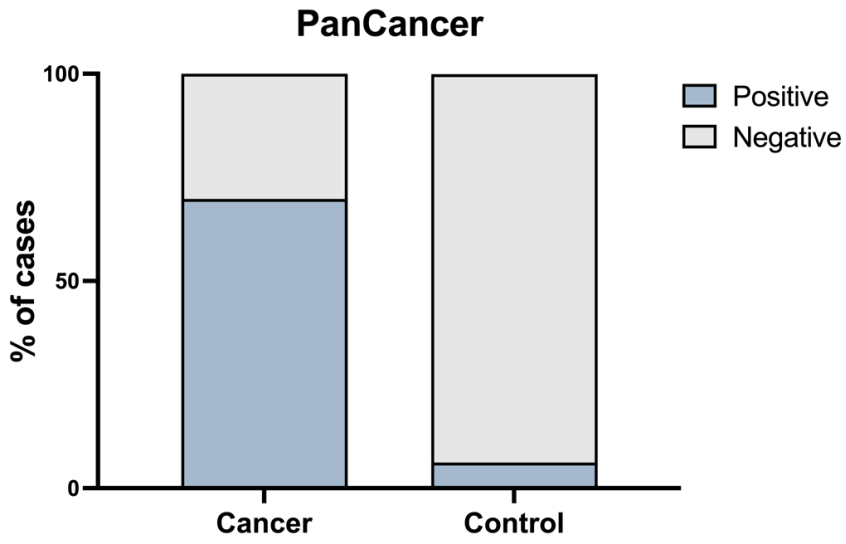

Figure 19. Percentage of cases identified by the "PanCancer" panel in cancer samples (70% Positive, 30% Negative) and in asymptomatic controls (6% Positive, 94% Negative).

5.4 Association between gene promoter methylation levels and clinicopathological features

Regarding associations between promoters' methylation levels and clinicopathological features, higher *MIR129-2* methylation levels were associated with advanced stages of LC ($p<0.01$; Figure 18A), CRC (Figure 20B-1) and PCa (Figure 20C-2). Additionally, higher $NID2_{me}$ levels were present in late-stage CRC (Figure 20B-2) and PCa (Figure 20C-3). $MAGI2_{me}$ and $ADCY4_{me}$ levels were also significantly higher in stage IV CRC and PCa, respectively, compared to earlier stages (Figure 20B-3 and C-1, respectively).

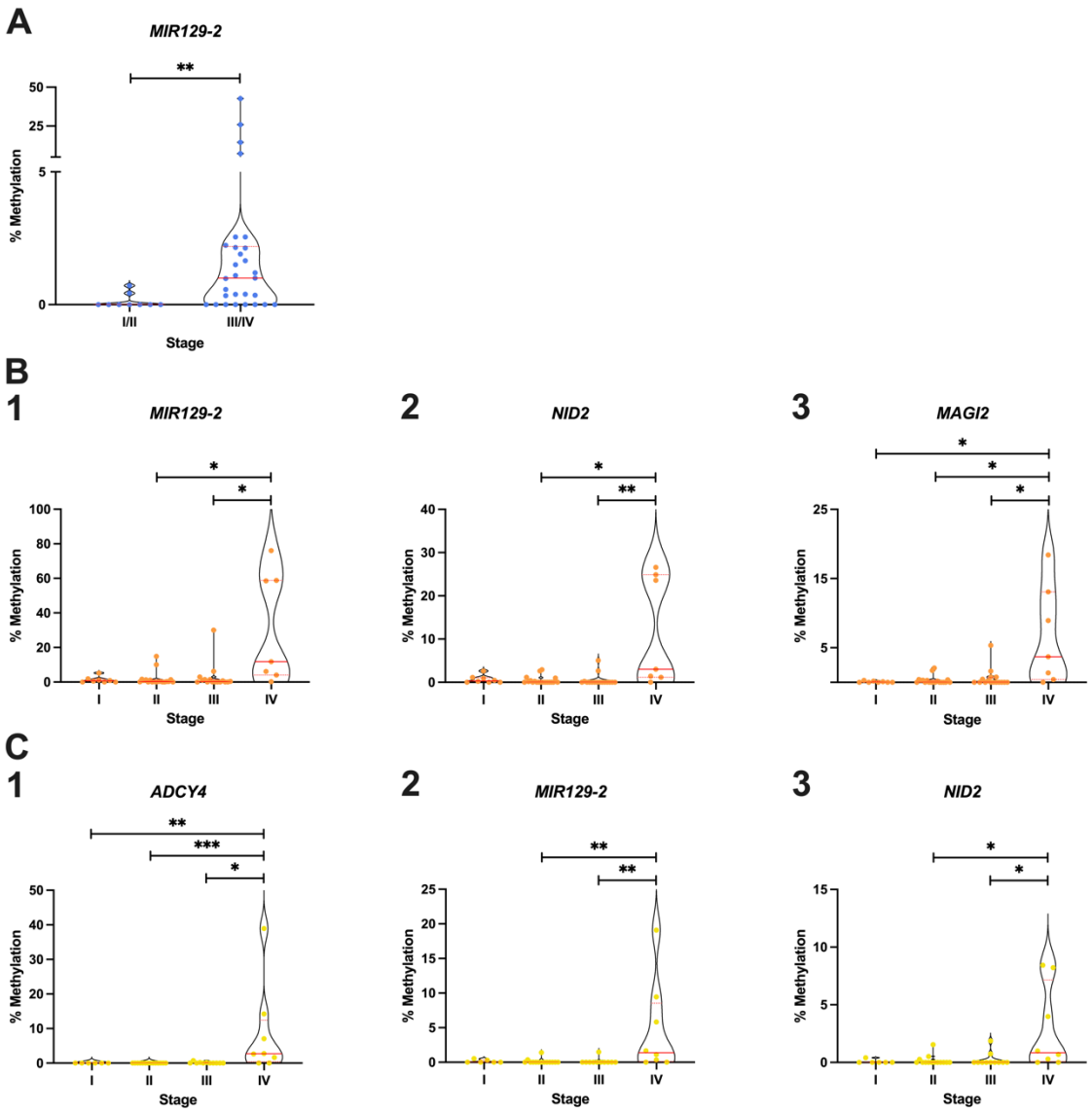

Figure 20. Distribution of methylation levels in lung (A), colorectal (B) and prostate (C) cancer patients according to clinical stage. (A) *MIR129-2* promoter methylation levels in stage I & II and III & IV lung cancer patients. (B)-(1) *MIR129-2*, (2) *NID2* and (3) *MAGI2* promoter's methylation levels across stage I-IV colorectal cancer patients. (C)-(1) *ADCY4*, (2) *MIR129-2* and (3) *NID2* promoter's methylation levels across stage I-IV prostate cancer patients. Kruskal-Wallis test followed by Dunn's multiple comparisons test, *p<0.05, **p<0.01, ***p<0.001. Red horizontal lines represent the median value and interquartile range.

Similarly, *MIR129-2*$_{me}$ higher levels were associated with metastatic disease (M) in LC (p=0.0150; Figure 21A), CRC (p<0.0001; Figure 21B-1) and PCa (p<0.01; Figure 21C-2) patients, while the same applies for *NID2*$_{me}$ levels in the 2 latter (Figure 21B-2 and C-3, respectively). *MAGI2*$_{me}$ levels were also significantly higher in CRC metastatic patients (p<0.0001; Figure 21B-3), whereas *ADCY4*$_{me}$ levels were more present in the circulation of PCa patients with distant metastasis (p<0.01; Figure 21C-1).

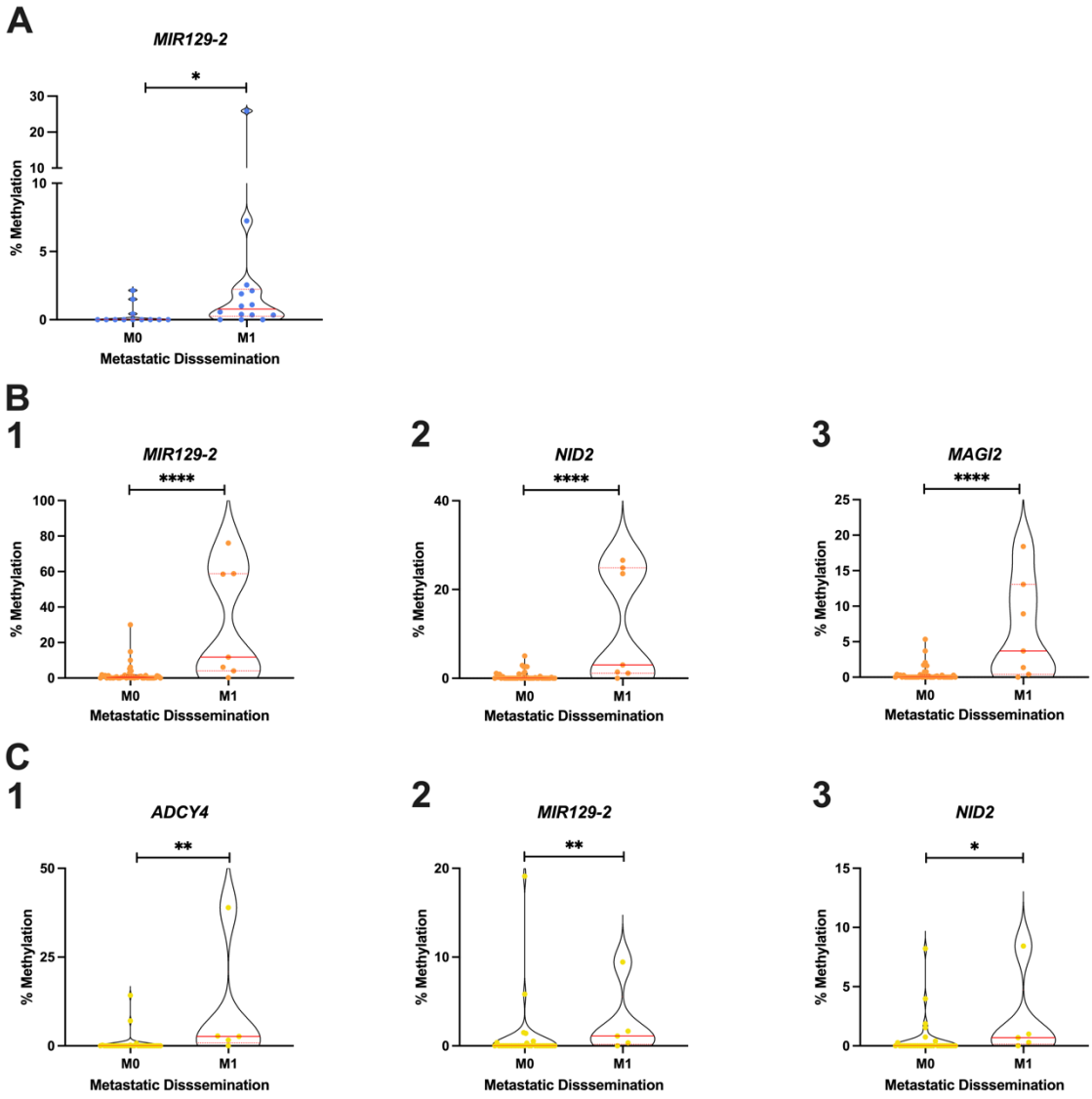

Figure 21. Distribution of methylation levels in lung (A), colorectal (B) and prostate (C) cancer patients according to metastatic dissemination. (A) *MIR129-2* promoter methylation levels in non-metastatic (M0) and metastatic (M+) lung cancer patients. (B)-(1) *MIR129-2*, (2) *NID2* and (3) *MAGI2* promoter's methylation levels across non-metastatic and metastatic cancer patients. (C)-(1) *ADCY4*, (2) *MIR129-2* and (3) *NID2* promoter's methylation levels across non-metastatic and metastatic prostate cancer patients. Mann-Whitney U Test, *p<0.05, **p<0.01, ****p<0.0001. Red horizontal lines represent the median value and interquartile range.

ADCY4, MIR129-2 and *NID2* promoter methylation levels were also significantly elevated in PCa patients with regional lymph node (N) involvement ($p<0.0001$ for $ADCY4_{me}$ and $MIR129\text{-}2_{me}$ and $p<0.001$ for $NID2_{me}$; Figure 22).

No associations were observed between clinicopathological features and circulating methylation levels in BrC samples.

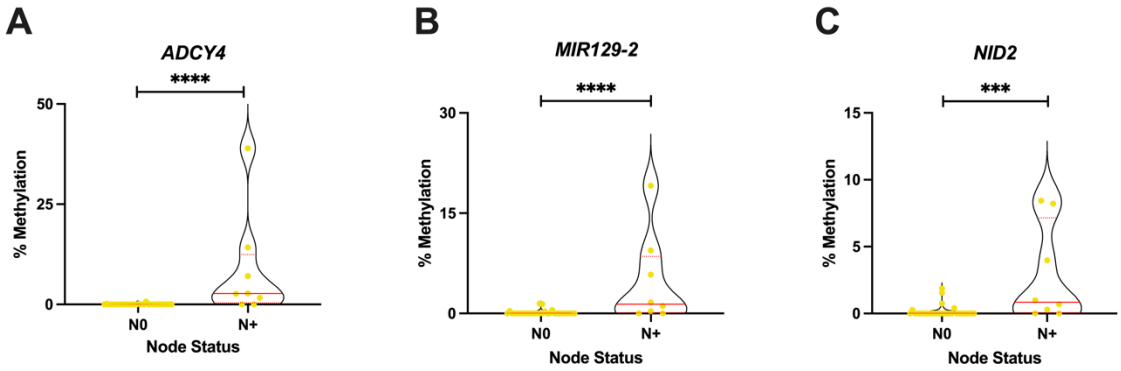

Figure 22. (A) *ADCY4*, (B) *MIR129-2* and (C) *NID2* promoter's methylation levels across node-positive (N+) and node-negative (N0) prostate cancer patients. Mann-Whitney U Test, ***p<0.001, ****p<0.0001. Red lines represent the median value and interquartile range.

VI.DISCUSSION

BrC, LC, CRC and PCa are the four most incident cancers worldwide and are among the deadliest, for both males and females, accounting for over 3 million annual deaths [1]. Although being part of the small fraction of cancers with available screening methods, which have, indeed, contributed to mortality reduction, several disadvantages, such as overdiagnosis [11, 17, 35] and invasiveness of the procedures [23] have also resulted from it. Thus, minimally invasive and highly specific screening protocols for early detection of these major cancers are urgently needed. DNA methylation displays cancer-specific aberrant patterns, namely hypermethylation and consequent silencing of tumor suppressor genes, in addition to being an early oncogenic event and easily accessible in cfDNA [66]. Besides, given its tissue-specific features, it shows great promise as a minimally-invasive biomarker for multi-cancer early detection using a single blood sample. Thus, we sought to assess the feasibility of a DNA methylation blood-based test for simultaneous detection of BrC, LC, CRC and PCa using multiplex ddPCR.

Aiming to identify candidate hypermethylated genes common to the four cancers, a TCGA data mining was performed, disclosing 30 potential genes. Additionally, cancer type-specific genes were also identified, 82 for BrC, 12 for LC, 496 for CRC and 159 for PCa. The top 5 genes common to all cancers were selected to comprise a “PanCancer” gene panel, while the top hypermethylated specific gene of each cancer was selected to compass a “CancerType” panel aiming to discriminate the TOO after a positive “PanCancer” result.

ADCY4, CDO1, MIR129-2, NID2 and *MAGI2* promoter methylation levels were compared *in silico* between tumoral and normal tissues of the four cancers using aggregation methylation values. Such analysis confirmed their potential as pan-cancer biomarkers, which we further validated using an in-house cohort of tissue samples. The methylation levels assessed by qMSP supported the *in silico* results, highlighting the potential value of these genes as markers for liquid-biopsy testing. Noteworthy, primers and probes could not be designed for *CDO1* fulfilling all our criteria, thus this gene was discarded since we could not ensure its methylation-specificity.

Adenylyl cyclase type 4 (*ADCY4*), like all adenylyl cyclases, is involved in the cAMP signaling pathway, which has been proposed as a treatment target, showing inhibition of cell growth and migration, as well as improvements in conventional antitumor drug sensitivity [140]. *ADCY4* functions as a tumor suppressor gene and has been shown, *in silico*, to be downregulated by promoter hypermethylation in BrC and LC. Besides, it has been proposed as a potential biomarker for the detection and prognosis of BrC [140, 141]. In the same line, in our study, higher methylation levels were found in CRC and PCa tumor samples compared to normal samples for this gene promotor.

MIR129-2 encodes the miR-129-2, which acts as a tumor suppressor microRNA in many cancers. Our results are in line with several others from other research groups, with

MIR129-2 promoter hypermethylation having been described in CRC cell lines and tissue samples [142], BrC primary tumors [141, 143-145], LC cell lines and primary tumors [146, 147] and PCa tissue specimens and urine sediments [148], supporting our results. Additionally, several other malignancies such as liquid tumors [149], gastric cancer [150], liver cancer [151, 152], bladder cancer [153], ovarian cancer [154], esophageal cancer [155], oropharyngeal carcinoma [156] and glioma [157] have been shown to display *MIR129-2* hypermethylation. Indeed, *MIR129-2* has been shown to be hypermethylated in multiple cancers, being considered one of the markers of the universal pan-cancer set [158].

Nidogen 2 (*NID2*) belongs to the nidogen protein family, thus being involved in maintaining the integrity and stability of basement membranes. *NID2* silencing by promoter hypermethylation has been described for LC in tissues [159], cell lines [160] and plasma samples [161], as well as for BrC tissues [162]. Other cancers like gastric [163], oral squamous cell carcinoma [164, 165] and bladder cancer [166-169] also displayed *NID2* hypermethylation. Besides LC and BrC, we observed high levels of *NID2* promoter hypermethylation for CRC and PCa.

MAGI2 encodes the membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2. This protein has been shown to be less expressed during PCa progression, suggesting that it can be helpful to predict PCa aggressiveness [170, 171]. Concerning *MAGI2* methylation, no reports are available concerning the tumor types address by our study, however, higher *MAGI2* methylation levels have been found in cervical cancer, both in tissue [172, 173] and cervical scrapings samples [174], as well as in ovarian [175] and gastric cancer primary tumors [176].

Overall, the available literature supports our hypothesis that the methylation levels of selected pan-cancer genes may indeed detect BrC, LC, CRC and PCa in a variety of biological samples and it further shows its extent to additional cancer types.

Regarding the “CancerType” selected genes, *HOXA11* promoter methylation levels were only significantly higher in LC tissues compared to respective normal, both *in silico* and in the tested tissue set, as expected. Nonetheless, high methylation levels were also seen in BrC and respective normal tissues, compromising its value as a cfDNA-based biomarker for LC discrimination, as both healthy and BrC females may, in theory, also display higher levels of circulating $HOXA11_{me}$. Homeobox A11 (*HOXA11*) belongs to the homeobox family of transcription factors, thus modulating gene expression in morphogenesis and differentiation. Its downregulation by promoter hypermethylation has been reported for LC and this gene is thought to have a tumor suppressor function, once it inhibited cell proliferation and migration in lung carcinogenesis in *in vitro* assays [177, 178]. Furthermore, in agreement with our results, *HOXA11* methylation has also been reported in tissue samples from BrC [179].

CELF2 was identified as the candidate gene for BrC discrimination and, although *in silico* analysis showed significantly higher methylation levels of the respective promoter in BrC, LC and PCa, by qMSP, significant differences were obtained only in BrC primary tumors. This gene encodes the CUGBP Elav-like family member 2 protein, an RNA-binding protein involved in pre-mRNA alternative splicing, mRNA translation and stability. Indeed, *CELF2* promoter hypermethylation associated with transcriptional silencing, enhancing breast tumors' growth [180]. Moreover, higher $CELF2_{me}$ circulating levels have also been reported in ovarian cancer [181].

Similarly, *in silico CHFR* promoter methylation levels were higher in LC and PCa in addition to CRC, while in our tissue set, significant differences were only depicted for CRC. Checkpoint with forkhead and ring finger domains (*CHFR*) encodes the E3 ubiquitin-protein ligase CHFR already identified as a mitotic stress checkpoint and tumor suppressor gene, that was shown to display promoter hypermethylation in CRC cell lines and tissue samples [182, 183]. However, many other cancer types have shown high methylation levels of *CHFR*, namely BrC [184], LC [185, 186], gastric cancer [187, 188], pancreatic cancer [189], esophageal cancer [190] and oral cancer [191].

FLOT1, identified as PCa specific by *in silico* analysis, also showed high methylation levels in normal breast tissues. Such gene encodes a scaffolding protein of lipid rafts, implicated in several cellular mechanisms, including signal transduction, protein recruiting, and cell proliferation. While FLOT1 overexpression has been associated with different cancer types, like LC [192, 193], BrC [194] and CRC [195], no data is available on *FLOT1* methylation.

Gathering our results with the available literature, the potential of this "CancerType" panel as a tool for discriminating each one of the tested cancers is rather weak, since higher methylation levels have been reported for many other cancers. Hence, additional genes were selected for each cancer from the data mining process.

EDNRB and *ZSCAN1* promoter methylation levels were significantly higher in our set of BrC, LC and CRC tissues, thus discarding their value as specific biomarkers for BrC and LC, respectively. Endothelin Receptor Type B (*EDNRB*) encodes a G-protein coupled receptor involved in phosphatidylinositol-calcium signaling and acts as a tumor suppressor gene in many cancers. Its downregulation by promoter hypermethylation has already been reported by others in CRC [196, 197], LC [198] and PCa [199-201], as well as in bladder cancer [202], gastric cancer [203], hepatocellular carcinoma [204] and head and neck cancer [205-207]. Regarding *ZSCAN1*, which encodes a zinc finger protein involved in DNA transcription, its methylation levels have only been reported for cervical cancer as a part of gene panels aiming to triage HPV-positive women [208, 209]. Nonetheless, in our hands, higher *GFRA1* promoter methylation levels were also found in CRC and LC. Although this

has been confirmed in LC [210], only *GFRA1* re-activation by hypomethylation has been associated with CRC aggressiveness, due to this protein being involved in the activation of the RET oncogene and downstream AKT signaling [211, 212]. Furthermore, we found higher *AOX1* promoter methylation levels in PCa and CRC samples. This gene encodes the aldehyde oxidase 1, an enzyme involved in oxidation-reduction reactions, that, in line with our data, was previously shown to be downregulated by methylation in PCa [213-215] and CRC [216].

Overall, the identified genes for TOO discrimination showed little value, suggesting that improvements are needed to develop a better algorithm for cancer-type-specific genes' selection. Thereby, we did not further assess their methylation levels in cfDNA and, instead, focused our efforts on developing a highly accurate ddPCR-based assay for detecting circulating methylation levels of the "PanCancer" genes.

Given the impact of pre-analytical conditions on cfDNA downstream analysis [217], we started by comparing different extraction methodologies across a set of plasma samples from cancer patients and healthy blood donors (optimization set). While the automatic extractor magLEAD® 12gC performed the poorest, the 2 manual kits showed similar performance concerning concentration values and cfDNA fragment size yield. Nonetheless, MagMAX™ Cell-Free DNA Isolation Kit outperformed QIAmp MinElute ccfDNA Kit in regard to contamination with high molecular weight DNA and %cfDNA. Besides, it provided a higher normalized cfDNA concentration, since the similar concentration values depicted between the 2 kits originated from different starting plasma volumes. Accordingly, the MagMAX™ kit was selected as the most suitable method for cfDNA extraction from plasma samples. Indeed, this kit has been shown to outperform other magnetic bead-based kits, providing the highest yield and low molecular weight fractions [218]. Other pre-analytical factors affecting cfDNA, such as the type of blood collection tubes, blood collection volume, plasma isolation, storage and freeze-thaw cycles could not be evaluated since the used samples were provided by the institutional biobank. Notwithstanding, blood was collected into EDTA tubes, processed within 4 hours after collection and plasma was carefully collected avoiding the buffy coat, aliquoted and frozen at -80ºC, thereby, following the most recommended guidelines [217].

In the optimization set samples, using a defined amount of 5 µL of DNA input per reaction, hypermethylation levels were only observed in samples of BrC, CRC and PCa patients with advanced stages of malignancy (stage IV) for all the gene promoters. Regarding the DNA input of each sample in the ddPCR reaction, a minimum of 5ng per reaction was mandatory in order to reliably detect methylation in cfDNA. Hence, such quantity was defined as a cut-off for sample eligibility. Additionally, another cut-off was further defined for the number of β-Actin copies per sample, aiming to control the cfDNA

input quantity and to minimize false-negative results. Following such reasoning, samples in the testing set were considered for methylation testing if a minimum of 10ng was available (5ng for each panel) and only analyzable if a minimum of 250 copies/µL of *ACTB* was reached.

After applying these criteria, from each group's 50 selected plasma samples, only 38 LC, 25 BrC, 47 CRC, 39 PCa and 33 AC samples were considered suitable for methylation analysis. This goes in agreement with the obtained cfDNA low concentration values for BrC and AC, thus leading to more excluded samples. In fact, insufficient cfDNA was a major drawback in this work, due to the limited amount of plasma available. Since we used biobank-stored samples, a maximum of 4mL of plasma per sample was available for cfDNA extraction. Comparing with other cfDNA methylation-based tests, both Epi ProColon® and Epi proLung® tests require 3.5 mL of plasma, nonetheless, these use a single gene panel, while our ddPCR assay consisted of 2, thus requiring the double input. The IvyGene test collects 40 mL of blood from each individual [219], while we routinely collect 12 mL (3 EDTA tubes of 4 mL) of blood from cancer patients at diagnosis and 8 mL (2 EDTA tubes of 4 mL) from blood donors. Thereby, the available amount of plasma for cfDNA extraction in the IvyGene test is far higher. Similarly, the Galleri® test requires collecting 20mL of blood per person [114], while in the CCGA study 80 mL were collected and cfDNA was extracted from 10mL of plasma [83]. Hence, up to 80mL of blood collection is feasible and safe, considering that healthy blood donors usually donate between 400-500 mL [220]. Hypothesizing a “real world” application of our test, higher blood volumes would be collected, and all resulting plasma used for DNA extraction, overcoming the lack of input challenge. Besides, even if such minimal input was not achieved, individuals would be contacted to re-collect a sample. Another approach would be to develop a 5-plex-based ddPCR assay combining all our targets, thus only requiring a limited sample to be inputted once, resulting in more eligibility. In fact, such high-order multiplexing has already been performed, although only for assessing gene expression [221, 222]. Also, new digital PCR technologies with more than 2 fluorescence channels have recently become available [223, 224].

Regarding the circulating methylation levels of “PanCancer” gene promoters, all displayed significantly higher levels in cancer patients compared to controls, except *MIR129-2* and *NID2* in BrC patients. To the best of our knowledge, we are the first team to report the methylation levels of these gene promoters in cfDNA of patients with LC, BrC, CRC and PCa. Only *NID2* methylation levels have been assessed in LC patients’ plasma, showing a 45.65% methylation rate [161]. Their performance in cancer detection was analyzed individually, displaying 100% specificity for $ADCY4_{me}$, $NID2_{me}$ and $MAGI2_{me}$ and 93.75% for $MIR129\text{-}2_{me}$ for all cancer types. Nonetheless, sensitivities fell short for BrC detection, with $NID2_{me}$ only being able to detect this cancer with 8%. In fact, $NID2_{me}$ showed

the lowest sensitivity for all cancer types, excepting PCa. Nonetheless, *MIR129-2*$_{me}$ was able to detect LC and CRC with 60.53% and 68.09% sensitivity, respectively. Aiming to increase detection sensitivity, we further constructed gene panels ("PanCancer" panel). When considering a sample positive whenever one of the genes was positive, sensitivity values increased to 52% for BrC, 85.71% for LC, 78.72% for CRC and 56.41% for PCa, at 93.75% specificity. 100% specificity was obtained when considering at least 2 positive genes, however, at the cost of decreased sensitivity. Indeed, our "PanCancer" panel may overcome a major challenge of standard screening procedures, with our high specificity resulting in a maximum false-positive rate of 6%, thus minimizing overdiagnosis and unnecessary costly additional workups. Although showing lower sensitivity than mammography, our test may be useful in detecting malignant lesions in women with dense breasts, often missed by the above mentioned [10], while also reducing radiation exposure by functioning as a tool for triaging women at higher risk of BrC. Similarly, our test also underperformed colonoscopy screening [24], but its minimally invasive nature may result in much higher patient compliance, hopefully encouraging individuals with a positive result to pursue a confirmatory colonoscopy. Notably, our results have clearly outperformed both LDCT [18] and PSA testing [225], thus our test may be a promising strategy for pre-screening LC and PCa and triaging patients for subsequent standard clinical approaches. Moreover, the minimal invasiveness, high specificity and low cost of our test favors its usage in an annual basis along with other routine blood tests, aiming to increase the detection sensitivity or even detect tumors that were too small to be detected by molecular analysis. Given the importance of detecting cancer at earlier stages, we further evaluated the "PanCancer" panel's ability for detecting stage I and II BrC, LC, CRC and PCa, displaying similar performance to all-stage detection.

In addition to individual cancer detection, we evaluated the capacity of the "PanCancer" panel to simultaneously detect BrC, LC, CRC and PCa, showing near 70% sensitivity at 93.75% specificity. While disclosing similar sensitivity to previous panels reported by our group for simultaneous detection of the major female [78] and male [77] cancer types, this panel showed much higher specificity. Nonetheless, these studies relied on whole-genome amplified cfDNA followed by qMSP analysis, while in the current study methylation-specific ddPCR, a more precise technology requiring less input DNA, was used [226]. Moreover, although our assay showed less sensitivity relatively to recently published multi-cancer early detection blood tests (Table 1), most of these are based on sequencing or other high-throughput methodologies, which are lengthy processes, require expert bioinformatic analysis and are highly costly [227, 228]. Conversely, we have developed a targeted, fast-workflow and cost-effective ddPCR assay, thus being more feasible as a population-based screening tool.

Importantly, gene promoter methylation levels associated with tumors clinicopathological features. For instance, circulating *MIR129-2*$_{me}$ levels were significantly elevated in late-stage and metastatic LC, CRC and PCa patients. Indeed, miR129-2 downregulation by methylation has been shown to render SOX4 active and promote CRC aggressiveness, including invasion and metastasis [229, 230]. Similarly, *NID2*$_{me}$ and *MAGI2*$_{me}$ levels were significantly higher in metastatic CRC, while higher *ADCY4*$_{me}$ and *NID2*$_{me}$ levels were found in node-positive and metastatic PCa at the time of diagnosis. Nonetheless, the number of samples from each clinical feature was very reduced, impairing the evaluation of the methylation levels of these gene promoters in distinguishing early- from late-stage cancer.

Overall, we report a proof-of-concept study showing the potential of a simple blood-based test for the simultaneous detection of the four most incident cancers worldwide. Nevertheless, it should be emphasized that we had a limited number of tested samples, in addition to a short follow-up of asymptomatic controls, thereby not being able to confirm that controls who tested positive will not eventually develop any cancer. Another drawback is the lack of a test for subsequent identification of tumor location, which hampers the choice of the most suitable confirmatory method. Despite that, by combining the gender, habits related to risk factors and family history of cancer of individuals with a positive “PanCancer” test, a more guided sequential workup may be possible. Additionally, our results should be validated in prospective studies including high-risk populations and benign conditions.

VII.CONCLUSION & FUTURE PERSPECTIVES

In conclusion, we confirm the feasibility of using a single blood test for detecting multiple cancers, particularly, BrC, LC, CRC and PCa, major incident and deadly cancers. Thus, such test can complement current screening methods, aiming to increase patient compliance and shifting cancer detection to earlier stages, where curative treatment options are more likely to succeed.

Considering our results, we further intend to:

- Develop a more suitable algorithm for identifying methylated genes with potential for discriminating between BrC, LC, CRC and PCa.
- Develop a 5-plex ddPCR assay for assessing the simultaneous promoter methylation levels of *ADCY4, MIR129-2, NID2* and *MAGI2*.
- Validate our results in a larger set of plasma samples from BrC, LC, CRC and PCa patients and healthy blood donors.
- Assess the methylation levels of *ADCY4, MIR129-2, NID2* and *MAGI2* in patients with benign lesions and additional cancer types.
- Validate our results in large-scale multicentre prospective studies, including high-risk populations.

VIII.REFERENCES

1. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.* CA: A Cancer Journal for Clinicians, 2021. **71**(3): p. 209-249.
2. Makki, J., *Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance.* Clin Med Insights Pathol, 2015. **8**: p. 23-31.
3. Rojas, K. and A. Stuckey, *Breast Cancer Epidemiology and Risk Factors.* Clin Obstet Gynecol, 2016. **59**(4): p. 651-672.
4. Barnard, M.E., C.E. Boeke, and R.M. Tamimi, *Established breast cancer risk factors and risk of intrinsic tumor subtypes.* Biochim Biophys Acta, 2015. **1856**(1): p. 73-85.
5. Eliyatkin, N., et al., *Molecular Classification of Breast Carcinoma: From Traditional, Old-Fashioned Way to A New Age, and A New Way.* J Breast Health, 2015. **11**(2): p. 59-66.
6. Harbeck, N. and M. Gnant, *Breast cancer.* The Lancet, 2017. **389**(10074): p. 1134-1150.
7. Niell, B.L., et al., *Screening for Breast Cancer.* Radiol Clin North Am, 2017. **55**(6): p. 1145-1162.
8. Senkus, E., et al., *Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2015. **26 Suppl 5**: p. v8-30.
9. Fiorica, J.V., *Breast Cancer Screening, Mammography, and Other Modalities.* Clin Obstet Gynecol, 2016. **59**(4): p. 688-709.
10. Yaffe, M.J., *Towards improving accuracy, effectiveness, and efficiency in breast cancer screening.* Lancet Oncol, 2018. **19**(11): p. 1426-1427.
11. Marmot, M.G., et al., *The benefits and harms of breast cancer screening: an independent review.* Br J Cancer, 2013. **108**(11): p. 2205-40.
12. Planchard, D., et al., *Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2018. **29**(Suppl 4): p. iv192-iv237.
13. Nasim, F., B.F. Sabath, and G.A. Eapen, *Lung Cancer.* Med Clin North Am, 2019. **103**(3): p. 463-473.
14. de Groot, P. and R.F. Munden, *Lung cancer epidemiology, risk factors, and prevention.* Radiol Clin North Am, 2012. **50**(5): p. 863-76.
15. Rodriguez-Canales, J., E. Parra-Cuentas, and Wistuba, II, *Diagnosis and Molecular Classification of Lung Cancer.* Cancer Treat Res, 2016. **170**: p. 25-46.
16. Bernhardt, E.B. and S.I. Jalal, *Small Cell Lung Cancer.* Cancer Treat Res, 2016. **170**: p. 301-22.
17. Wu, G.X. and D.J. Raz, *Lung Cancer Screening.* Cancer Treat Res, 2016. **170**: p. 1-23.
18. National Lung Screening Trial Research, T., et al., *Reduced lung-cancer mortality with low-dose computed tomographic screening.* N Engl J Med, 2011. **365**(5): p. 395-409.
19. Colson, Y.L., J.O. Shepard, and I.T. Lennes, *New USPSTF Guidelines for Lung Cancer Screening: Better but Not Enough.* JAMA Surg, 2021. **156**(6): p. 513-514.
20. Brenner, H., M. Kloor, and C.P. Pox, *Colorectal cancer.* Lancet, 2014. **383**(9927): p. 1490-1502.
21. Aran, V., et al., *Colorectal Cancer: Epidemiology, Disease Mechanisms and Interventions to Reduce Onset and Mortality.* Clin Colorectal Cancer, 2016. **15**(3): p. 195-203.
22. Molnar, B., et al., *Plasma methylated septin 9: a colorectal cancer screening marker.* Expert Rev Mol Diagn, 2015. **15**(2): p. 171-84.
23. Simon, K., *Colorectal cancer development and advances in screening.* Clinical interventions in aging, 2016. **11**: p. 967-976.
24. Geneve, N., et al., *Colorectal Cancer Screening.* Prim Care, 2019. **46**(1): p. 135-148.
25. Bailey, J.R., A. Aggarwal, and T.F. Imperiale, *Colorectal Cancer Screening: Stool DNA and Other Noninvasive Modalities.* Gut Liver, 2016. **10**(2): p. 204-11.

26. Labianca, R., et al., *Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2013. **24 Suppl 6**: p. vi64-72.
27. Perdana, N.R., et al., *The Risk Factors of Prostate Cancer and Its Prevention: A Literature Review.* Acta Med Indones, 2016. **48**(3): p. 228-238.
28. Attard, G., et al., *Prostate cancer.* Lancet, 2016. **387**(10013): p. 70-82.
29. Humphrey, P.A., *Histopathology of Prostate Cancer.* Cold Spring Harb Perspect Med, 2017. **7**(10).
30. Wang, G., et al., *Genetics and biology of prostate cancer.* Genes Dev, 2018. **32**(17-18): p. 1105-1140.
31. Arora, K. and C.E. Barbieri, *Molecular Subtypes of Prostate Cancer.* Curr Oncol Rep, 2018. **20**(8): p. 58.
32. Roobol, M.J. and S.V. Carlsson, *Risk stratification in prostate cancer screening.* Nat Rev Urol, 2013. **10**(1): p. 38-48.
33. Hoffman, R.M., *Clinical practice. Screening for prostate cancer.* N Engl J Med, 2011. **365**(21): p. 2013-9.
34. Barry, M.J. and L.H. Simmons, *Prevention of Prostate Cancer Morbidity and Mortality: Primary Prevention and Early Detection.* Med Clin North Am, 2017. **101**(4): p. 787-806.
35. Etzioni, R., et al., *Overdiagnosis Due to Prostate-Specific Antigen Screening: Lessons From U.S. Prostate Cancer Incidence Trends.* JNCI: Journal of the National Cancer Institute, 2002. **94**(13): p. 981-990.
36. Mottet, N., et al., *EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent.* Eur Urol, 2017. **71**(4): p. 618-629.
37. World Health, O. and S. International Programme on Chemical, *Biomarkers in risk assessment : validity and validation*. 2001, World Health Organization: Geneva.
38. Costa-Pinheiro, P., et al., *Diagnostic and prognostic epigenetic biomarkers in cancer.* Epigenomics, 2015. **7**(6): p. 1003-15.
39. Wu, L. and X. Qu, *Cancer biomarker detection: recent achievements and challenges.* Chemical Society Reviews, 2015. **44**(10): p. 2963-2997.
40. Goossens, N., et al., *Cancer biomarker discovery and validation.* Transl Cancer Res, 2015. **4**(3): p. 256-269.
41. Cheng, F., L. Su, and C. Qian, *Circulating tumor DNA: a promising biomarker in the liquid biopsy of cancer.* Oncotarget, 2016. **7**(30): p. 48832-48841.
42. Poulet, G., J. Massias, and V. Taly, *Liquid Biopsy: General Concepts.* Acta Cytol, 2019. **63**(6): p. 449-455.
43. Marrugo-Ramírez, J., M. Mir, and J. Samitier, *Blood-Based Cancer Biomarkers in Liquid Biopsy: A Promising Non-Invasive Alternative to Tissue Biopsy.* Int J Mol Sci, 2018. **19**(10).
44. Ignatiadis, M., G.W. Sledge, and S.S. Jeffrey, *Liquid biopsy enters the clinic — implementation issues and future challenges.* Nature Reviews Clinical Oncology, 2021. **18**(5): p. 297-312.
45. Peng, X., et al., *Identifying the tissues-of-origin of circulating cell-free DNAs is a promising way in noninvasive diagnostics.* Brief Bioinform, 2021. **22**(3).
46. Schwarzenbach, H., D.S. Hoon, and K. Pantel, *Cell-free nucleic acids as biomarkers in cancer patients.* Nat Rev Cancer, 2011. **11**(6): p. 426-37.
47. Han, X., J. Wang, and Y. Sun, *Circulating Tumor DNA as Biomarkers for Cancer Detection.* Genomics Proteomics Bioinformatics, 2017. **15**(2): p. 59-72.
48. von Bubnoff, N., *Liquid Biopsy: Approaches to Dynamic Genotyping in Cancer.* Oncology Research and Treatment, 2017. **40**(7-8): p. 409-416.
49. Keller, L., et al., *Clinical relevance of blood-based ctDNA analysis: mutation detection and beyond.* British Journal of Cancer, 2021. **124**(2): p. 345-358.
50. Siravegna, G., et al., *Integrating liquid biopsies into the management of cancer.* Nature Reviews Clinical Oncology, 2017. **14**(9): p. 531-548.

51. Snow, A., D. Chen, and J.E. Lang, *The current status of the clinical utility of liquid biopsies in cancer.* Expert Rev Mol Diagn, 2019. **19**(11): p. 1031-1041.
52. US Food and Drug Administration. *Guardant360 CDx – P200010/S001*. [September 2022]; Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf20/P200010S001C.pdf.
53. US Food and Drug Administration. *FoundationOne Liquid CDx – P200016*. [September 2022]; Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf20/P200016C.pdf.
54. Bettegowda, C., et al., *Detection of circulating tumor DNA in early- and late-stage human malignancies.* Sci Transl Med, 2014. **6**(224): p. 224ra24.
55. Alix-Panabières, C. and K. Pantel, *Liquid Biopsy: From Discovery to Clinical Application.* Cancer Discovery, 2021. **11**(4): p. 858-873.
56. Rodríguez-Paredes, M. and M. Esteller, *Cancer epigenetics reaches mainstream oncology.* Nature Medicine, 2011. **17**(3): p. 330-339.
57. Jeronimo, C. and R. Henrique, *Epigenetic biomarkers in urological tumors: A systematic review.* Cancer Lett, 2014. **342**(2): p. 264-74.
58. Sharma, S., T.K. Kelly, and P.A. Jones, *Epigenetics in cancer.* Carcinogenesis, 2010. **31**(1): p. 27-36.
59. Inbar-Feigenberg, M., et al., *Basic concepts of epigenetics.* Fertil Steril, 2013. **99**(3): p. 607-15.
60. Werner, R.J., A.D. Kelly, and J.J. Issa, *Epigenetics and Precision Oncology.* Cancer J, 2017. **23**(5): p. 262-269.
61. Robertson, K.D., *DNA methylation, methyltransferases, and cancer.* Oncogene, 2001. **20**(24): p. 3139-55.
62. Hatziapostolou, M. and D. Iliopoulos, *Epigenetic aberrations during oncogenesis.* Cell Mol Life Sci, 2011. **68**(10): p. 1681-702.
63. Pan, Y., et al., *DNA methylation profiles in cancer diagnosis and therapeutics.* Clin Exp Med, 2018. **18**(1): p. 1-14.
64. Baylin, S.B. and P.A. Jones, *Epigenetic Determinants of Cancer.* Cold Spring Harbor perspectives in biology, 2016. **8**(9): p. a019505.
65. Sweet, T.J. and A.H. Ting, *WOMEN IN CANCER THEMATIC REVIEW: Diverse functions of DNA methylation: implications for prostate cancer and beyond.* Endocr Relat Cancer, 2016. **23**(11): p. T169-T178.
66. Constâncio, V., et al., *DNA Methylation-Based Testing in Liquid Biopsies as Detection and Prognostic Biomarkers for the Four Major Cancer Types.* Cells, 2020. **9**(3): p. 624.
67. Beltran-Garcia, J., et al., *Epigenetic IVD Tests for Personalized Precision Medicine in Cancer.* Front Genet, 2019. **10**: p. 621.
68. Roy, D. and M. Tiirikainen, *Diagnostic Power of DNA Methylation Classifiers for Early Detection of Cancer.* Trends Cancer, 2020. **6**(2): p. 78-81.
69. Lamb, Y.N. and S. Dhillon, *Epi proColon((R)) 2.0 CE: A Blood-Based Screening Test for Colorectal Cancer.* Mol Diagn Ther, 2017. **21**(2): p. 225-232.
70. Clarke, C.A., et al., *Projected Reductions in Absolute Cancer-Related Deaths from Diagnosing Cancers Before Metastasis, 2006-2015.* Cancer Epidemiol Biomarkers Prev, 2020. **29**(5): p. 895-902.
71. Ahlquist, D.A., *Universal cancer screening: revolutionary, rational, and realizable.* npj Precision Oncology, 2018. **2**(1): p. 23.
72. Dong, S., et al., *Histone-Related Genes Are Hypermethylated in Lung Cancer and Hypermethylated HIST1H4F Could Serve as a Pan-Cancer Biomarker.* Cancer Res, 2019. **79**(24): p. 6101-6112.
73. Moskalev, E.A., et al., *GHSR DNA hypermethylation is a common epigenetic alteration of high diagnostic value in a broad spectrum of cancers.* Oncotarget, 2014. **6**(6).

74. Dai, X., et al., *Site-Specific Hypermethylation of SST 1stExon as a Biomarker for Predicting the Risk of Gastrointestinal Tract Cancers.* Dis Markers, 2022. **2022**: p. 4570290.
75. Margolin, G., et al., *Robust Detection of DNA Hypermethylation of ZNF154 as a Pan-Cancer Locus with in Silico Modeling for Blood-Based Diagnostic Development.* J Mol Diagn, 2016. **18**(2): p. 283-98.
76. Miller, B.F., H.M. Petrykowska, and L. Elnitski, *Assessing ZNF154 methylation in patient plasma as a multicancer marker in liquid biopsies from colon, liver, ovarian and pancreatic cancer patients.* Sci Rep, 2021. **11**(1): p. 221.
77. Constâncio, V., et al., *Early detection of the major male cancer types in blood-based liquid biopsies using a DNA methylation panel.* Clin Epigenetics, 2019. **11**(1): p. 175.
78. Nunes, S.P., et al., *Cell-Free DNA Methylation of Selected Genes Allows for Early Detection of the Major Cancers in Women.* Cancers (Basel), 2018. **10**(10).
79. Tan, S.H., et al., *Detection of promoter hypermethylation in serum samples of cancer patients by methylation-specific polymerase chain reaction for tumour suppressor genes including RUNX3.* Oncol Rep, 2007. **18**(5): p. 1225-30.
80. Melson, J., et al., *Commonality and differences of methylation signatures in the plasma of patients with pancreatic cancer and colorectal cancer.* International Journal of Cancer, 2014. **134**(11): p. 2656-2662.
81. Kang, S., et al., *CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA.* Genome Biology, 2017. **18**(1): p. 53.
82. Li, W., et al., *CancerDetector: ultrasensitive and non-invasive cancer detection at the resolution of individual reads using cell-free DNA methylation sequencing data.* Nucleic Acids Res, 2018. **46**(15): p. e89.
83. Liu, M.C., et al., *Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA.* Annals of Oncology, 2020. **31**(6): p. 745-759.
84. Klein, E.A., et al., *Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set.* Ann Oncol, 2021. **32**(9): p. 1167-1177.
85. Kandimalla, R., et al., *EpiPanGI Dx: A Cell-free DNA Methylation Fingerprint for the Early Detection of Gastrointestinal Cancers.* Clin Cancer Res, 2021. **27**(22): p. 6135-6144.
86. Chen, X., et al., *Non-invasive early detection of cancer four years before conventional diagnosis using a blood test.* Nat Commun, 2020. **11**(1): p. 3475.
87. Shen, S.Y., et al., *Sensitive tumour detection and classification using plasma cell-free DNA methylomes.* Nature, 2018. **563**(7732): p. 579-583.
88. Liu, L., et al., *Targeted methylation sequencing of plasma cell-free DNA for cancer detection and classification.* Ann Oncol, 2018. **29**(6): p. 1445-1453.
89. IvyGene Labs. *The IvyGeneCORE Test*. [August, 2022]; Available from: https://www.ivygenelabs.co.za/the-ivygenecore-test/
90. Gao, Q., et al., *LBA3 Early detection and localization of multiple cancers using a blood-based methylation assay (ELSA-seq).* Annals of Oncology, 2020. **31**: p. S1358.
91. Han, T., et al., *An ultrasensitive method for noninvasive pan-cancer early detection based on targeted methylation sequencing of cell-free DNA.* Journal of Clinical Oncology, 2021. **39**(15_suppl): p. 10544-10544.
92. Han, T., et al., *An ultrasensitive approach for cancer screening and tissue of origin prediction based on targeted methylation sequencing of cell-free DNA.* Journal of Clinical Oncology, 2022. **40**(16_suppl): p. 10553-10553.
93. Kinross, J., et al., *97P A panel of methylation markers for multi-cancer detection from plasma.* Annals of Oncology, 2020. **31**: p. S280.

94. Valouev, A., et al., *Abstract 2141: Development of a highly-sensitive targeted cell-free DNA epigenomic assay for early-stage multi-cancer screening.* Cancer Research, 2022. **82**(12_Supplement): p. 2141-2141.
95. Zhao, G.Q., et al., *A novel NGS kit solution for multi-cancer early detection using circulating cell free DNA based methylation analysis.* Journal of Clinical Oncology, 2022. **40**(16_suppl): p. 10542-10542.
96. Xu, L., et al., *Abstract 4601: Toward the development of a $100 screening test for 6 major cancer types.* Cancer Research, 2020. **80**(16_Supplement): p. 4601-4601.
97. Xu, L., et al., *Abstract 2610: A high performance blood test for multiple cancer early screening.* Cancer Research, 2021. **81**(13_Supplement): p. 2610-2610.
98. Xu, L., et al., *Validation of a high performing blood test for multiple major cancer screenings.* Journal of Clinical Oncology, 2021. **39**(15_suppl): p. 10561-10561.
99. Sina, A.A.I., et al., *Epigenetically reprogrammed methylation landscape drives the DNA self-assembly and serves as a universal cancer biomarker.* Nature Communications, 2018. **9**(1): p. 4915.
100. Nagasaka, T., et al., *Analysis of Fecal DNA Methylation to Detect Gastrointestinal Neoplasia.* JNCI: Journal of the National Cancer Institute, 2009. **101**(18): p. 1244-1258.
101. Katerov, S., et al., *Abstract 111: Accurate multi-cancer detection using methylated DNA markers and proteins in plasma.* Cancer Research, 2021. **81**(13_Supplement): p. 111-111.
102. Allawi, H.T., et al., *Abstract 631: Validation of a panel of methylated DNA and protein markers for multi-cancer detection in plasma.* Cancer Research, 2022. **82**(12_Supplement): p. 631-631.
103. Chan, K.C.A., et al., *Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing.* Proceedings of the National Academy of Sciences, 2013. **110**(47): p. 18761-18768.
104. Stackpole, M., et al., *Abstract 24: Multi-feature ensemble learning on cell-free dna for accurately detecting and locating cancer.* Cancer Research, 2021. **81**(13_Supplement): p. 24-24.
105. Li, W., et al., *5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers.* Cell Res, 2017. **27**(10): p. 1243-1257.
106. Bergamaschi, A., et al., *Pilot study demonstrating changes in DNA hydroxymethylation enable detection of multiple cancers in plasma cell-free DNA.* medRxiv, 2020: p. 2020.01.22.20018382.
107. Hao, X., et al., *DNA methylation markers for diagnosis and prognosis of common cancers.* Proceedings of the National Academy of Sciences, 2017. **114**(28): p. 7414-7419.
108. Singlera Genomics. *Singlera Genomics' PanSeer assay technology*. [August 2022]; Available from: https://singleraoncology.com/technology/.
109. Sheridan, C., *Grail to pour $1 billion into blood test to detect early cancer.* Nat Biotechnol, 2017. **35**(2): p. 101-102.
110. Fiala, C. and E.P. Diamandis, *Can Grail find the trail to early cancer detection?* Clin Chem Lab Med, 2019. **57**(4): p. 403-406.
111. Klein, E.A., et al., *Development of a comprehensive cell-free DNA (cfDNA) assay for early detection of multiple tumor types: The Circulating Cell-free Genome Atlas (CCGA) study.* Journal of Clinical Oncology, 2018. **36**(15_suppl): p. 12021-12021.
112. Liu, M.C., et al., *Plasma cell-free DNA (cfDNA) assays for early multi-cancer detection: The circulating cell-free genome atlas (CCGA) study.* Annals of Oncology, 2018. **29**: p. viii14.
113. Chen, X., et al., *Prognostic Significance of Blood-Based Multi-cancer Detection in Plasma Cell-Free DNA.* Clin Cancer Res, 2021. **27**(15): p. 4221-4229.
114. GRAIL. *The Galleri® test*. [September 2022]; Available from: https://www.galleri.com/the-galleri-test/request-the-test.

115. Nadauld, L.D., et al., *The PATHFINDER Study: Assessment of the Implementation of an Investigational Multi-Cancer Early Detection Test into Clinical Practice.* Cancers, 2021. **13**(14): p. 3501.
116. Beer, T.M., et al., *Interim results of PATHFINDER, a clinical use study using a methylation-based multi-cancer early detection test.* Journal of Clinical Oncology, 2021. **39**(15_suppl): p. 3010-3010.
117. Kint, S., et al., *Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR.* PLoS One, 2018. **13**(6): p. e0199091.
118. Sørensen, A.L. and P. Collas, *Immunoprecipitation of methylated DNA.* Methods Mol Biol, 2009. **567**: p. 249-62.
119. Adela Inc. *A breakthrough approach to early detection*. [August 2022]; Available from: https://www.adelabio.com/.
120. Shen, S.Y., et al., *Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA.* Nature Protocols, 2019. **14**(10): p. 2749-2780.
121. Szyf, M., *The elusive role of 5'-hydroxymethylcytosine.* Epigenomics, 2016. **8**(11): p. 1539-1551.
122. Bluestar Genomics. *Multi-cancer Panels*. [August 2022]; Available from: https://www.bluestargenomics.com/multi-cancer-panels/.
123. Zou, H., et al., *T2036 Pan-Detection of Gastrointestinal Neoplasms By Stool DNA Testing: Establishment of Feasibility.* Gastroenterology, 2009. **136**(5): p. A-625.
124. Yang, D., et al., *Patient perceptions of stool DNA testing for pan-digestive cancer screening: a survey questionnaire.* World J Gastroenterol, 2014. **20**(17): p. 4972-9.
125. Cohen, J.D., et al., *Detection and localization of surgically resectable cancers with a multi-analyte blood test.* Science, 2018. **359**(6378): p. 926-930.
126. Lennon, A.M., et al., *Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention.* Science, 2020. **369**(6499).
127. van der Pol, Y. and F. Mouliere, *Toward the Early Detection of Cancer by Decoding the Epigenetic and Environmental Fingerprints of Cell-Free DNA.* Cancer Cell, 2019. **36**(4): p. 350-368.
128. Snyder, M.W., et al., *Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin.* Cell, 2016. **164**(1-2): p. 57-68.
129. DELFI Diagnostics. *Our Science*. [August 2022]; Available from: https://delfidiagnostics.com/our-science/.
130. Cristiano, S., et al., *Genome-wide cell-free DNA fragmentation in patients with cancer.* Nature, 2019. **570**(7761): p. 385-389.
131. Li, Y., D. Ge, and C. Lu, *The SMART App: an interactive web application for comprehensive DNA methylation analysis and visualization.* Epigenetics & Chromatin, 2019. **12**(1): p. 71.
132. Agilent Technologies. *Performance Characteristics of the Cell-Free DNA ScreenTape Assay*. 2019.
133. Patterson, K., et al., *DNA methylation: bisulphite modification and analysis.* J Vis Exp, 2011(56).
134. Armbruster, D.A. and T. Pry, *Limit of blank, limit of detection and limit of quantitation.* Clin Biochem Rev, 2008. **29 Suppl 1**(Suppl 1): p. S49-52.
135. Milbury, C.A., et al., *Determining lower limits of detection of digital PCR assays for cancer-related gene mutations.* Biomolecular Detection and Quantification, 2014. **1**(1): p. 8-22.
136. Yu, M., T.J. Heinzerling, and W.M. Grady, *DNA Methylation Analysis Using Droplet Digital PCR.* Methods Mol Biol, 2018. **1768**: p. 363-383.
137. Schisterman, E.F., et al., *Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples.* Epidemiology, 2005. **16**(1): p. 73-81.
138. Whale, A.S., J.F. Huggett, and S. Tzonev, *Fundamentals of multiplexing with digital PCR.* Biomol Detect Quantif, 2016. **10**: p. 15-23.

139. Saelee, S.L., et al., *Quantitative PCR-Based Method to Assess Cell-Free DNA Quality, Adjust Input Mass, and Improve Next-Generation Sequencing Assay Performance.* The Journal of Molecular Diagnostics, 2022. **24**(6): p. 566-575.
140. Fan, Y., et al., *Epigenetic identification of ADCY4 as a biomarker for breast cancer: an integrated analysis of adenylate cyclases.* Epigenomics, 2019. **11**(14): p. 1561-1579.
141. Mao, X.H., et al., *Identification of differentially methylated genes as diagnostic and prognostic biomarkers of breast cancer.* World J Surg Oncol, 2021. **19**(1): p. 29.
142. Bandres, E., et al., *Epigenetic regulation of microRNA expression in colorectal cancer.* Int J Cancer, 2009. **125**(11): p. 2737-43.
143. Burdennyy, A.M., et al., *Optimized Marker System for Early Diagnosis of Breast Cancer.* Bull Exp Biol Med, 2021. **172**(1): p. 57-62.
144. Filippova, E.A., et al., *Relationship of the Levels of microRNA Gene Methylation with the Level of Their Expression and Pathomorphological Characteristics of Breast Cancer.* Bull Exp Biol Med, 2021. **171**(6): p. 764-769.
145. Braga, E.A., et al., *Marker Systems Based on MicroRNA Gene Methylation for the Diagnosis of Stage I-II Breast Cancer.* Bull Exp Biol Med, 2019. **168**(2): p. 280-284.
146. Xiao, Y., et al., *Epigenetic regulation of miR-129-2 and its effects on the proliferation and invasion in lung cancer cells.* J Cell Mol Med, 2015. **19**(9): p. 2172-80.
147. Heller, G., et al., *DNA methylation of microRNA-coding genes in non-small-cell lung cancer patients.* J Pathol, 2018. **245**(4): p. 387-398.
148. Torres-Ferreira, J., et al., *MiR-193b promoter methylation accurately detects prostate cancer in urine sediments and miR-34b/c or miR-129-2 promoter methylation define subsets of clinically aggressive tumors.* Mol Cancer, 2017. **16**(1): p. 26.
149. Wong, K.Y., et al., *Epigenetic inactivation of the MIR129-2 in hematological malignancies.* J Hematol Oncol, 2013. **6**: p. 16.
150. Alizadeh, N., et al., *Evaluation of the Methylation of MIR129-2 Gene in Gastric Cancer.* J Gastrointest Cancer, 2020. **51**(1): p. 267-270.
151. Chen, X., et al., *Methylation-mediated repression of microRNA 129-2 enhances oncogenic SOX4 expression in HCC.* Liver Int, 2013. **33**(3): p. 476-86.
152. Lu, C.Y., et al., *Frequent DNA methylation of MiR-129-2 and its potential clinical implication in hepatocellular carcinoma.* Genes Chromosomes Cancer, 2013. **52**(7): p. 636-43.
153. Padrão, N.A., et al., *MicroRNA promoter methylation: a new tool for accurate detection of urothelial carcinoma.* Br J Cancer, 2017. **116**(5): p. 634-639.
154. Loginov, V.I., et al., *Aberrant Methylation of 20 miRNA Genes Specifically Involved in Various Steps of Ovarian Carcinoma Spread: From Primary Tumors to Peritoneal Macroscopic Metastases.* Int J Mol Sci, 2022. **23**(3).
155. Chen, X., et al., *CpG island methylation status of miRNAs in esophageal squamous cell carcinoma.* Int J Cancer, 2012. **130**(7): p. 1607-13.
156. Ren, S., et al., *Discovery and development of differentially methylated regions in human papillomavirus-related oropharyngeal squamous cell carcinoma.* Int J Cancer, 2018. **143**(10): p. 2425-2436.
157. Yang, Y., et al., *MiR-129-2 functions as a tumor suppressor in glioma cells by targeting HMGB1 and is down-regulated by DNA methylation.* Mol Cell Biochem, 2015. **404**(1-2): p. 229-39.
158. Vrba, L. and B.W. Futscher, *A suite of DNA methylation markers that can detect most common human cancers.* Epigenetics, 2018. **13**(1): p. 61-72.
159. Geng, J., et al., *Methylation status of NEUROG2 and NID2 improves the diagnosis of stage I NSCLC.* Oncol Lett, 2012. **3**(4): p. 901-906.
160. Wang, J., et al., *Silencing NID2 by DNA Hypermethylation Promotes Lung Cancer.* Pathol Oncol Res, 2020. **26**(2): p. 801-811.

161. Feng, X., et al., *The more potential performance of nidogen 2 methylation by tissue or plasma DNA over brichoalveolar lavage DNA in diagnosis of nonsmall cell lung cancer.* J Cancer Res Ther, 2018. **14**(Supplement): p. S341-S346.
162. Strelnikov, V.V., et al., *Abnormal promoter DNA hypermethylation of the integrin, nidogen, and dystroglycan genes in breast cancer.* Sci Rep, 2021. **11**(1): p. 2264.
163. Ulazzi, L., et al., *Nidogen 1 and 2 gene promoters are aberrantly methylated in human gastrointestinal cancer.* Mol Cancer, 2007. **6**: p. 17.
164. Guerrero-Preston, R., et al., *NID2 and HOXA9 promoter hypermethylation as biomarkers for prevention and early detection in oral cavity squamous cell carcinoma tissues and saliva.* Cancer Prev Res (Phila), 2011. **4**(7): p. 1061-72.
165. Srisuttee, R., et al., *Evaluation of NID2 promoter methylation for screening of Oral squamous cell carcinoma.* BMC Cancer, 2020. **20**(1): p. 218.
166. Renard, I., et al., *Identification and validation of the methylated TWIST1 and NID2 genes through real-time methylation-specific polymerase chain reaction assays for the noninvasive detection of primary bladder cancer in urine samples.* Eur Urol, 2010. **58**(1): p. 96-104.
167. El Azzouzi, M., et al., *Evaluation of DNA methylation in promoter regions of hTERT, TWIST1, VIM and NID2 genes in Moroccan bladder cancer patients.* Cancer Genet, 2022. **260-261**: p. 41-45.
168. Hermanns, T., et al., *A noninvasive urine-based methylation biomarker panel to detect bladder cancer and discriminate cancer grade.* Urol Oncol, 2020. **38**(6): p. 603.e1-603.e7.
169. Fantony, J.J., et al., *Urinary NID2 and TWIST1 methylation to augment conventional urine cytology for the detection of bladder cancer.* Cancer Biomark, 2017. **18**(4): p. 381-387.
170. David, S.N., et al., *MAGI2 is an independent predictor of biochemical recurrence in prostate cancer.* Prostate, 2018. **78**(8): p. 616-622.
171. Mahdian, R., et al., *Expression profile of MAGI2 gene as a novel biomarker in combination with major deregulated genes in prostate cancer.* Mol Biol Rep, 2014. **41**(9): p. 6125-31.
172. Loaeza-Loaeza, J., et al., *The CpG island methylator phenotype increases the risk of high-grade squamous intraepithelial lesions and cervical cancer.* Clin Epigenetics, 2022. **14**(1): p. 4.
173. Chen, Y.C., et al., *Methylomics analysis identifies epigenetically silenced genes and implies an activation of β-catenin signaling in cervical cancer.* Int J Cancer, 2014. **135**(1): p. 117-27.
174. Chang, C.C., et al., *The feasibility of detecting endometrial and ovarian cancer using DNA methylation biomarkers in cervical scrapings.* J Gynecol Oncol, 2018. **29**(1): p. e17.
175. Valle, B.L., et al., *HIST1H2BB and MAGI2 Methylation and Somatic Mutations as Precision Medicine Biomarkers for Diagnosis and Prognosis of High-grade Serous Ovarian Cancer.* Cancer Prev Res (Phila), 2020. **13**(9): p. 783-794.
176. Qu, Y., N. Gao, and T. Wu, *Expression and clinical significance of SYNE1 and MAGI2 gene promoter methylation in gastric cancer.* Medicine (Baltimore), 2021. **100**(4): p. e23788.
177. Hwang, J.A., et al., *HOXA11 hypermethylation is associated with progression of non-small cell lung cancer.* Oncotarget, 2013. **4**(12): p. 2317-25.
178. Li, Q., et al., *DNA methylation profiling identifies the HOXA11 gene as an early diagnostic and prognostic molecular marker in human lung adenocarcinoma.* Oncotarget, 2017. **8**(20): p. 33100-33109.
179. Shan, M., et al., *DNA Methylation Profiles and Their Diagnostic Utility in BC.* Dis Markers, 2019. **2019**: p. 6328503.
180. Pique, L., et al., *Epigenetic inactivation of the splicing RNA-binding protein CELF2 in human breast cancer.* Oncogene, 2019. **38**(45): p. 7106-7112.

181. Marinelli, L.M., et al., *Methylated DNA markers for plasma detection of ovarian cancer: Discovery, validation, and clinical feasibility.* Gynecol Oncol, 2022. **165**(3): p. 568-576.
182. Toyota, M., et al., *Epigenetic inactivation of CHFR in human tumors.* Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7818-23.
183. Sun, Z., et al., *The diagnostic and prognostic value of CHFR hypermethylation in colorectal cancer, a meta-analysis and literature review.* Oncotarget, 2017. **8**(51): p. 89142-89148.
184. Murria, R., et al., *Methylation of tumor suppressor genes is related with copy number aberrations in breast cancer.* Am J Cancer Res, 2015. **5**(1): p. 375-85.
185. Guo, M., et al., *CHFR methylation strongly correlates with methylation of DNA damage repair and apoptotic pathway genes in non-small cell lung cancer.* Discov Med, 2015. **19**(104): p. 151-8.
186. Wang, C., et al., *Clinicopathological significance of CHFR methylation in non-small cell lung cancer: a systematic review and meta-analysis.* Oncotarget, 2017. **8**(65): p. 109732-109739.
187. Li, Y., et al., *Predictive value of CHFR and MLH1 methylation in human gastric cancer.* Gastric Cancer, 2015. **18**(2): p. 280-7.
188. Dai, D., et al., *CHFR Promoter Hypermethylation Is Associated with Gastric Cancer and Plays a Protective Role in Gastric Cancer Process.* J Cancer, 2019. **10**(4): p. 949-956.
189. Maekawa, H., et al., *Analysis of the methylation of CpG islands in the CDO1, TAC1 and CHFR genes in pancreatic ductal cancer.* Oncol Lett, 2020. **19**(3): p. 2197-2204.
190. Mei, X., et al., *The hypermethylation of the CDKN2A and CHFR promoter region is a key regulatory mechanism of CDKN2A and CHFR expression in esophageal squamous cell carcinoma.* Transl Cancer Res, 2019. **8**(3): p. 770-778.
191. Mori, K., et al., *Detecting Early-Stage Oral Cancer from Clinically Diagnosed Oral Potentially Malignant Disorders by DNA Methylation Profile.* Cancers (Basel), 2022. **14**(11).
192. Zhang, L., et al., *FLOT1 promotes tumor development, induces epithelial-mesenchymal transition, and modulates the cell cycle by regulating the Erk/Akt signaling pathway in lung adenocarcinoma.* Thorac Cancer, 2019. **10**(4): p. 909-917.
193. Li, H., et al., *Abnormal expression of FLOT1 correlates with tumor progression and poor survival in patients with non-small cell lung cancer.* Tumour Biol, 2014. **35**(4): p. 3311-5.
194. Li, L., et al., *Microrna-124 targets flotillin-1 to regulate proliferation and migration in breast cancer.* Mol Cancer, 2013. **12**: p. 163.
195. Baig, N., et al., *Clinical significance and comparison of flotillin 1 expression in left and right colon cancer.* Oncol Lett, 2019. **18**(2): p. 997-1004.
196. Mousavi Ardehaie, R., et al., *Aberrant methylated EDNRB can act as a potential diagnostic biomarker in sporadic colorectal cancer while KISS1 is controversial.* Bioengineered, 2017. **8**(5): p. 555-564.
197. Chen, C., et al., *Hypermethylation of EDNRB promoter contributes to the risk of colorectal cancer.* Diagn Pathol, 2013. **8**: p. 199.
198. Chen, S.C., et al., *Aberrant promoter methylation of EDNRB in lung cancer in Taiwan.* Oncol Rep, 2006. **15**(1): p. 167-72.
199. Yuan, Y., et al., *The value of endothelin receptor type B promoter methylation as a biomarker for the risk assessment and diagnosis of prostate cancer: A meta-analysis.* Pathol Res Pract, 2020. **216**(2): p. 152796.
200. Rogers, C.G., et al., *High concordance of gene methylation in post-digital rectal examination and post-biopsy urine samples for prostate cancer detection.* J Urol, 2006. **176**(5): p. 2280-4.
201. Jerónimo, C., et al., *Endothelin B receptor gene hypermethylation in prostate adenocarcinoma.* J Clin Pathol, 2003. **56**(1): p. 52-5.

202. Zuiverloon, T.C., et al., *A methylation assay for the detection of non-muscle-invasive bladder cancer (NMIBC) recurrences in voided urine.* BJU Int, 2012. **109**(6): p. 941-8.
203. Tao, K., et al., *Quantitative analysis of promoter methylation of the EDNRB gene in gastric cancer.* Med Oncol, 2012. **29**(1): p. 107-12.
204. Hsu, L.S., et al., *Aberrant methylation of EDNRB and p16 genes in hepatocellular carcinoma (HCC) in Taiwan.* Oncol Rep, 2006. **15**(2): p. 507-11.
205. Mydlarz, W.K., et al., *Serum biomarkers for detection of head and neck squamous cell carcinoma.* Head Neck, 2016. **38**(1): p. 9-14.
206. Shen, S., et al., *Targeting Viral DNA and Promoter Hypermethylation in Salivary Rinses for Recurrent HPV-Positive Oropharyngeal Cancer.* Otolaryngol Head Neck Surg, 2020. **162**(4): p. 512-519.
207. Fung, S.Y.H., 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**(6): p. 1812-1822.
208. van Leeuwen, R.W., et al., *DNA methylation markers as a triage test for identification of cervical lesions in a high risk human papillomavirus positive screening cohort.* Int J Cancer, 2019. **144**(4): p. 746-754.
209. Li, N., et al., *DNA methylation markers as triage test for the early identification of cervical lesions in a Chinese population.* Int J Cancer, 2021. **148**(7): p. 1768-1777.
210. Koh, Y.W., et al., *Association between the CpG island methylator phenotype and its prognostic significance in primary pulmonary adenocarcinoma.* Tumour Biol, 2016. **37**(8): p. 10675-84.
211. Dong, Z., et al., *Hypomethylation of GDNF family receptor alpha 1 promotes epithelial-mesenchymal transition and predicts metastasis of colorectal cancer.* PLoS Genet, 2020. **16**(11): p. e1009159.
212. Ma, W.R., et al., *Impact of GFRA1 gene reactivation by DNA demethylation on prognosis of patients with metastatic colon cancer.* World J Gastroenterol, 2020. **26**(2): p. 184-198.
213. Geybels, M.S., et al., *Epigenomic profiling of DNA methylation in paired prostate cancer versus adjacent benign tissue.* Prostate, 2015. **75**(16): p. 1941-50.
214. Li, Y., et al., *Diagnosis and prognosis potential of four gene promoter hypermethylation in prostate cancer.* Cell Biol Int, 2021. **45**(1): p. 117-126.
215. Wu, J., et al., *DNA Methylation-Mediated Lowly Expressed AOX1 Promotes Cell Migration and Invasion of Prostate Cancer.* Urol Int, 2022: p. 1-9.
216. Oster, B., et al., *Identification and validation of highly frequent CpG island hypermethylation in colorectal adenomas and carcinomas.* Int J Cancer, 2011. **129**(12): p. 2855-66.
217. Meddeb, R., E. Pisareva, and A.R. Thierry, *Guidelines for the Preanalytical Conditions for Analyzing Circulating Cell-Free DNA.* Clin Chem, 2019. **65**(5): p. 623-633.
218. Markus, H., et al., *Evaluation of pre-analytical factors affecting plasma DNA analysis.* Sci Rep, 2018. **8**(1): p. 7375.
219. IvyGene Labs. *Frequently Asked Questions.* [cited [October 2022; Available from: https://www.ivygenelabs.co.za/frequently-asked-questions/.
220. Van Remoortel, H., et al., *The effect of a standard whole blood donation on oxygen uptake and exercise capacity: a systematic review and meta-analysis.* Transfusion, 2017. **57**(2): p. 451-462.
221. Zavridou, M., et al., *Development and Analytical Validation of a 6-Plex Reverse Transcription Droplet Digital PCR Assay for the Absolute Quantification of Prostate Cancer Biomarkers in Circulating Tumor Cells of Patients with Metastatic Castration-Resistant Prostate Cancer.* Clin Chem, 2022. **68**(10): p. 1323-1335.
222. Strati, A., et al., *Development and Analytical Validation of a One-Step Five-Plex RT-ddPCR Assay for the Quantification of SARS-CoV-2 Transcripts in Clinical Samples.* Analytical Chemistry, 2022. **94**(36): p. 12314-12322.

223. Bio-Rad Laboratories, I. *QX600 Droplet Digital PCR System*. [October 2022]; Available from: https://www.bio-rad.com/en-pt/product/qx600-droplet-digital-pcr-system?ID=b07d12ac-0585-fc4c-a586-3ddf20d5c4a0.
224. QIAGEN. *QIAcuity dPCR products*. [October 2022]; Available from: https://www.qiagen.com/us/applications/digital-pcr/products.
225. Kearns, J.T. and D.W. Lin, *Improving the Specificity of PSA Screening with Serum and Urine Markers.* Curr Urol Rep, 2018. **19**(10): p. 80.
226. Pinheiro, L. and K.R. Emslie, *Basic Concepts and Validation of Digital PCR Measurements.* Methods Mol Biol, 2018. **1768**: p. 11-24.
227. Ewalt, M.D., H. West, and D.L. Aisner, *Next Generation Sequencing—Testing Multiple Genetic Markers at Once.* JAMA Oncology, 2019. **5**(7): p. 1076-1076.
228. Soto, J., et al., *The impact of next-generation sequencing on the DNA methylation-based translational cancer research.* Transl Res, 2016. **169**: p. 1-18.e1.
229. Chen, Z., et al., *MicroRNA-129 inhibits colorectal cancer cell proliferation, invasion and epithelial-to-mesenchymal transition by targeting SOX4.* Oncol Rep, 2021. **45**(5).
230. Rezayi Soufiani, A., et al., *Hypermethylation of MIR129-2 Regulates SOX4 Transcription and Associates with Metastasis in Patients with Colorectal Cancer.* J Gastrointest Cancer, 2022. **53**(3): p. 718-724.

IX.APPENDIX

I. Review article manuscript in preparation for submission to a peer-reviewed journal

Shifting the cancer screening paradigm: the rising potential of blood-based multi-cancer early detection tests

Tiago Brito-Rocha [1,2], Vera Constâncio [1,3], Rui Henrique [1,4,5] and Carmen Jerónimo [1,5]

[1] Cancer Biology and Epigenetics Group, Research Center of IPO Porto (CI-IPOP)/ RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO-Porto)/ Porto Comprehensive Cancer Centre (Porto.CCC)

[2] Master Programme in Oncology, School of Medicine & Biomedical Sciences, University of Porto (ICBAS-UP)

[3] Doctoral Programme in Biomedical Sciences, School of Medicine & Biomedical Sciences, University of Porto (ICBAS-UP)

[4] Department of Pathology, Portuguese Oncology Institute of Porto (IPOP)

[5] Department of Pathology and Molecular Immunology, School of Medicine & Biomedical Sciences, University of Porto (ICBAS-UP)

Abstract: Cancer remains a leading cause of death worldwide, partly due to being detected in late stages, where treatment options are scarce. Most cancers do not have recommended screening procedures, and the available ones present several drawbacks, leading to low patient compliance and unnecessary workups, adding up the costs to health care systems. Thus, there is a great need for accurate and minimally invasive tools for cancer early detection. In recent years, multi-cancer early detection (MCED) tests emerged as the ideal tool, combining molecular analysis of tumor-related markers present in body fluids with artificial intelligence to simultaneously detect a variety of cancers and further discriminate the underlying cancer type. Therefore, the aim of this review is to highlight the variety of different approaches currently being developed for MCED, in addition to the factors precluding their clinical implementation. Although showing great potential, large clinical validation of MCED tests is still lacking.