

RESEARCH

Open Access



COL25A1 and METAP1D DNA methylation are promising liquid biopsy epigenetic biomarkers of colorectal cancer using digital PCR

Alexis Overs^{1,2*}, Paul Peixoto², Eric Hervouet², Chloé Molimard³, Franck Monnien³, Jules Durand², Michael Guittaut², Angélique Vienot⁴, Julien Viot⁴, Michael Herfs⁵, Christophe Borg^{2,4}, Jean-Paul Feugeas^{1,2} and Zohair Selmani^{1,2}

Abstract

Background Colorectal cancer is a public health issue and was the third leading cause of cancer-related death worldwide in 2022. Early diagnosis can improve prognosis, making screening a central part of colorectal cancer management. Blood-based screening, diagnosis and follow-up of colorectal cancer patients are possible with the study of cell-free circulating tumor DNA. This study aimed to identify novel DNA methylation biomarkers of colorectal cancer that can be used for the follow-up of patients with colorectal cancer.

Methods A DNA methylation profile was established in the Gene Expression Omnibus (GEO) database ($n = 507$) using bioinformatics analysis and subsequently confirmed using The Cancer Genome Atlas (TCGA) database ($n = 348$). The in silico profile was then validated on local tissue and cell-free DNA samples using methylation-specific digital PCR in colorectal cancer patients ($n = 35$) and healthy donors ($n = 35$).

Results The DNA methylation of *COL25A1* and *METAP1D* was predicted to be a colorectal cancer biomarker by bioinformatics analysis (ROC AUC = 1, 95% CI [0.999–1]). The two biomarkers were confirmed with tissue samples, and the combination of *COL25A1* and *METAP1D* yielded 49% sensitivity and 100% specificity for cell-free DNA.

Conclusion Bioinformatics analysis of public databases revealed *COL25A1* and *METAP1D* DNA methylation as clinically applicable liquid biopsies DNA methylation biomarkers. The specificity implies an excellent positive predictive value for follow-up, and the high sensitivity and relative noninvasiveness of a blood-based test make these biomarkers compatible with colorectal cancer screening. However, the clinical impact of these biomarkers in colorectal cancer screening and follow-up needs to be established in further prospective studies.

Keywords COL25A1, METAP1D, DNA methylation, Liquid biopsy, Circulating tumor DNA, Biomarker, Colorectal cancer, Digital PCR

*Correspondence:

Alexis Overs

aovers@chu-besancon.fr

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Colorectal cancer (CRC) is a major public health concern worldwide, and it is the second leading cause of cancer-related mortality in 2022 and is a leading cause of morbidity and mortality [1]. Early detection of CRC allows for local endoscopic treatment. In contrast, late diagnosis of advanced or metastatic CRC requires more severe and systemic treatments with increased morbidity or incurable disease [2]. Therefore, CRC screening in the general population has been shown to reduce mortality [3]. Stool-based immunoassays is the most frequently used method for CRC screening, with a sensitivity of 74% and a specificity of 94%. However, the positive predictive value is only approximately 9% in the general population [4]. A positive screening test is usually followed by a confirmatory colonoscopy, many of which are negative. The number of unnecessary invasive procedures could be reduced with a more efficient screening test, highlighting the need for a highly sensitive and specific screening test. The required sensitivity and specificity could be achieved with the tumor DNA in liquid biopsy. Liquid biopsy can also be used to follow monitor CRC patients, in addition to the radiologic and clinical follow-up.

Tumor mutations can be detected in circulating tumor DNA (ctDNA) and are being increasingly used in clinical practice for CRC [5]. The detection of CRC mutations in ctDNA is routinely used in clinical practice to detect *KRAS*, *NRAS* and *BRAF* hotspot mutations and for patient follow-up [6]. However, only approximately half of CRC patients harbor a *KRAS/NRAS* hotspot [7].

Most cancer phenotypes can be explained by epigenetic modifications that can occur early in cancer development [8]. DNA methylation is an epigenetic mechanism that results in the addition of a methyl group to the fifth carbon of a cytosine. Global hypomethylation associated with hypermethylation of tumor suppressor genes is observed in most cancers [9]. New CRC epigenetic biomarkers are being developed for clinical practice, with the Food and Drug Administration (FDA) approval of EpiProcolon[®] (blood-based using *SEPT9* methylation) and Cologuard[®] (stool-based using a combination of *KRAS* hotspot mutation, fecal hemoglobin, *NDRG4* and *BMP3* methylation) [10]. The Shield[®] ctDNA panel has also been tested on a large scale for CRC screening with promising results [11]. Blood-based biomarkers can be more easily standardized for collection and analysis than stool-based tests, and they prevent the stool manipulation, which may limit the adoption of such tests by the population. These new blood biomarkers of ctDNA require highly sensitive technologies such as digital droplet PCR (ddPCR) [12] or deep sequencing. At present, the cost of deep sequencing makes it unlikely that large methylation profiles (up to hundreds of biomarkers)

could be implemented on a large scale, in contrast to ddPCR technologies. Therefore, our efforts are focused on a limited methylation profile (one to six biomarkers) that can be explored by ddPCR technologies using multiplexing.

The noninvasiveness, lack of stool manipulation and low cost of blood-based ddPCR may allow the test to be repeated over the years. The test needs to be highly sensitive for screening and highly specific to increase the positive predictive value and decrease the number of false positives.

Initially based on bioinformatics analysis, the aim of this study was to develop new DNA methylation CRC biomarkers that can be detected in ctDNA using ddPCR to develop a noninvasive blood-based CRC test with an efficiency compatible with clinical practice for the follow-up of CRC patients.

Methods

Study design

This study is based on a bioinformatics analysis that predicts potential biomarkers in the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. Then, the biomarkers were confirmed in tissue and plasma samples from the University Hospital of Besancon and the University of Liege.

Databases

Methylation data were downloaded from two independent databases, GEO and TCGA. In the GEO database, the GSE48684 ($n=147$), GSE129364 ($n=72$), GSE32149 ($n=73$) and GSE42921 ($n=23$) datasets were used for tumor and non-tumor colorectal tissue samples. The GSE62992 ($n=100$) and GSE36054 ($n=192$) series were also used for whole blood samples from healthy subjects. For TCGA, the TCGA-COAD dataset ($n=455$), which contains non-tumor colorectal tissue and CRC samples, was used. All the data were generated on the Illumina Infinium HumanMethylation450k Beadchip[®], which analyzes more than 450,000 CpGs across the genome, covering 96% of the CpG islands in the human genome.

Bioinformatics analysis

General design

The analyses were performed using a script written in R 3.5.1 [13] and the packages pROC 1.15.3 [14], GEOquery 2.56.0 [15] and TCGAbiolink 2.16.4 [16]. The main steps of the analysis were the normalization of the data, the selection of the differentially methylated CpG, the design of a panel and the comparison of the performance of the panel on another database.

Normalization

For each CpG, the B value was the ratio of methylated alleles to the total number of alleles. B values had a bimodal distribution between 0 and 1 and were heteroscedastic at extreme values. Therefore, the B values were not suitable for normalization and had to be converted to M values with $M = \log_2 \left(\frac{\beta}{1-\beta} \right)$. M values were homoscedastic even for highly methylated or unmethylated CpG and were more suitable for normalization transformations [17]. Normalization was performed independently for positive and negative M values with equalization of the medians between the series. The M values were then converted to B values using the following formula $\beta = \frac{2^m}{2^m+1}$.

Selection

A three-step selection of the CpG in HumanMethylation450k was performed. First, the CpGs were grouped into clusters by genomic proximity below 2000 base pairs, which is the size of a CpG island, and only the most discriminating CpG of each cluster between blood and CRC samples was retained. Second, the differentially methylated CpGs between colorectal samples and whole blood samples were selected. Finally, the differentially methylated CpGs between CRC and non-tumor colorectal tissue samples were selected.

Panel design

The area under the ROC curve (AUC) was used to construct the best-performing panel of CpGs. An iterative algorithm was performed, with each step including in the panel the CpG that most increased the AUC. The panel size was chosen by the optimal performance with the minimum number of CpGs, considering the possibility of probes for which quantitative methylation-specific PCR (qMSP) primers cannot be designed. The panel was then compared with the TCGA-COAD dataset. The AUC confidence intervals and performance differences were confirmed using a two-sided Delong test [18].

Panel on other tissues

The panel was tested on TCGA breast, pancreatic, lung, stomach, liver, esophagus, and skin tumor and non-tumor samples ($n = 893, 195, 919, 397, 472, 200$ and 475 , respectively).

Transcriptional impact

The transcriptional association of CpG methylation associated with genes was evaluated on the TCGA-COAD dataset. Transcriptional data were normalized using

Deseq2 [19], and gene expression was compared between non-tumor and tumor samples using two-sided Student's t tests.

Panel score value

In the in silico analysis and qMSP assay, the panel score value represented the mean B value of the panel. In the ddPCR assay, the panel score value represented the mean of the methylated biomarker copy number.

Experimental confirmation

Samples

A cohort of 19 formalin-fixed, paraffin-embedded (FFPE) CRC tissue samples and 15 non-tumor colorectal tissue samples was obtained from the University of Liège, Belgium.

The number of plasma sample was calculated using the epiR package [20] to estimate the sensitivity and specificity of the biomarkers at 90% and a 10% absolute confidence interval with in a cohort with a 50% prevalence.

The cohort of plasma samples consisted of 35 CRC patients from the University Hospital of Besancon and 35 healthy donors from the Etablissement Français du Sang (Table 1).

The sample inclusion criteria were patients followed at the University Hospital of Besancon with a pathological diagnosis, stage III-IV between 2022 and 2024.

For both CRC patients and healthy donors, 20 mL of blood was collected from a peripheral vein with two Cell-Free DNA BCT[®] (Streck, USA), followed by plasma purification. The purification process was a two-step centrifugation, 1300 g for 10 min at 4 °C, and a second centrifugation 13000 g for 10 min at 4 °C, after each centrifugation the plasma was manually separated from the cell pellet. The plasma was then stored at -20 °C until DNA extraction.

DNA extraction and conversion

DNA was extracted from FFPE samples using the QIAmp DNA Mini Kit[®] (Qiagen, Netherlands), and plasma DNA was extracted from 4 mL of plasma using the QIAmp Circulating Nucleic Acid Kit[®] (Qiagen, Netherlands) according to the manufacturer's instructions. DNA quantification was performed using a Qubit[®] fluorometer (Invitrogen, USA) directly after DNA extraction, and the DNA was kept in the QIAmp elution buffer at -20 °C until bisulfite conversion. DNA from FFPE and plasma samples was bisulfited and converted using the Diagenode Premium Bisulfite Kit[®] according to the manufacturer's instructions. After extraction, the DNA was either directly used for the experiments or stored at -20 °C.

Table 1 Clinical characteristics of patients and healthy donors' plasma samples

	Plasma from patients with CRC <i>n</i> = 35	Plasma from healthy donors <i>n</i> = 35	<i>p</i> value
Age, median (range)	70 (41–80)	45 (19–68)	$p < 10E-9$, Wilcoxon
Sex, <i>n</i> (%)			
Male	22 (31.4)	21 (30)	$p = 1$, χ^2
Female	13 (18.6)	13 (18.6)	
Not specified	0 (0)	1 (1.4)	
Type			
Adenocarcinoma NOS (<i>n</i> , %)	35 (100)	–	
Stage, <i>n</i> (%)			
III	10 (29)	–	
IV	25 (71)	–	
Not available	0 (0)	–	
Molecular status, <i>n</i> (%)			
Wild type	13 (37)	–	
KRAS mutated	12 (34)	–	
NRAS mutated	2 (6)	–	
BRAF mutated	2 (6)	–	
Not available	6 (17)	–	
Microsatellite status, <i>n</i> (%)			
MSS	31 (89)	–	
MSI	1 (3)	–	
Not available	3 (9)	–	

qMSP

To confirm the presence of differentially methylated CpG, qMSP was performed on converted DNA from FFPE samples. qMSP was performed in the same condition for methylated and unmethylated primers using Takara MasterMix® (Takara, Japan) on a StepOne® (Applied Biosystems, USA) thermal cycler with the following program: 95 °C for 600 s, 40 cycles of 95 °C for 60 s and 65 °C for 50 s. The qMSP primers were designed using MethPrimer [21] and BiSearch [22]. Amplification specificity was confirmed by gel electrophoresis and sequencing. Sanger sequencing was performed for amplification products above 120 base pairs, and pyrosequencing was performed for products below 120 base pairs. The *B* value was calculated using standard curves generated with the EpiTect PCR Control DNA Set® (Qiagen, USA).

ddPCR

Plasma methylation was performed using the Bio-Rad QX200 ddPCR® with Bio-Rad EVAGreen ddPCR Supermix® and Bio-Rad ddPCR Supermix for probes®. The level of methylation was assessed by the methylated biomarker copy number per microliter calculated from the number of positive and negative droplets using Bio-Rad QuantaSoft® software. The positivity threshold was set at 7500 LU in the FAM channel for *METAP1D* and 7000

LU for *COL25A1* and C1-Cless in the EVAGreen channel. A C-Less marker was used as a DNA control [23]. The C-Less marker was a PCR target that did not contain cytosine in its DNA sequence. The PCR amplification of such markers was unaffected by DNA methylation and bisulfite conversion processes. The length of the C-Less amplicon was 69 bp, which was comparable to that of the new biomarkers and effectively reflected the amount of analyzable DNA in the samples. The samples with no biomarker and a C-less than 250 copies per microliter were considered inconclusive due to the lack of cell-free DNA.

Results

Bioinformatics

The bioinformatics analysis performed on the GEO database revealed a panel of six CpGs from the 485,577 region. The selection process consisted of filtering and panel construction based on the methylation mean of the CpGs and their association in a binary classification model to classify non-tumor colorectal tissue versus CRC (Fig. 1).

In the first and second steps, the CpGs were clustered by genomic proximity, and 375,963 CpGs that were methylated in the blood or not methylated in CRC were excluded because these CpGs would not contribute to the blood-based assay. In the third step, the 100 most

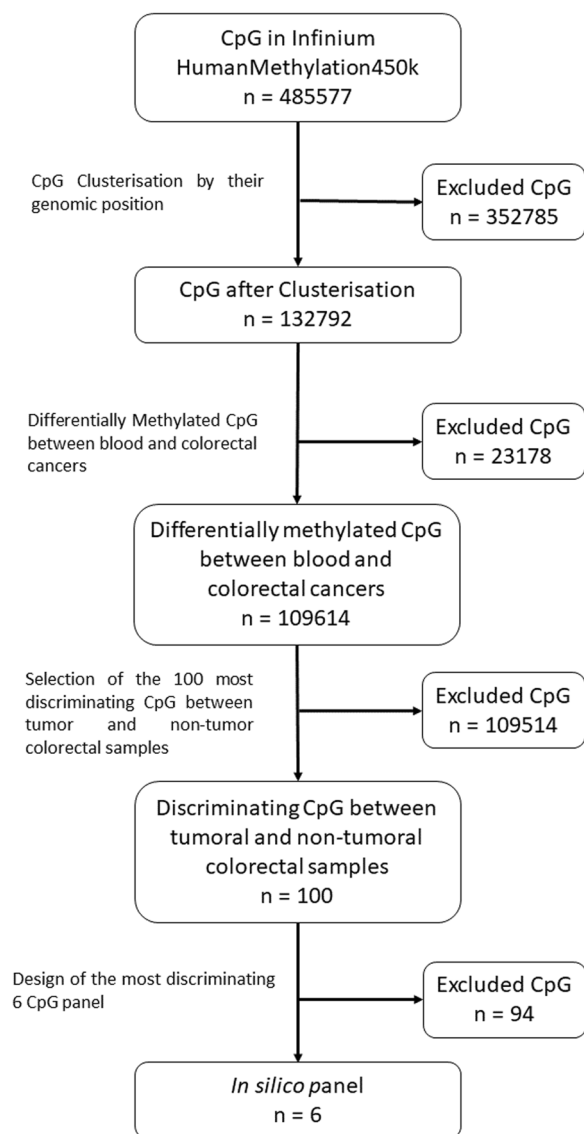


Fig. 1 Flowchart of the CpG selection from Infinium HumanMethylation450k to the panel. CpG was selected with three filters to keep the 100 most discriminating between blood, tumor and non-tumor colorectal samples. The panel was the most discriminating combination of 6 CpG among the 100 selected

differentially methylated CpGs between CRC and non-tumor tissues were selected. During the panel construction, the panel performance increased from one to two CpGs (AUC = 0.986 vs 0.996), and the addition of another CpG did not increase panel performance (Fig. 2).

Interestingly, all the CpGs in the panel had similar differential methylation between CRC and non-tumor colorectal tissues (Supplementary data 1).

The generated panel consisted of a combination of six CpGs located on chromosomes 1, 2, 3, 4 and 8. Four of these CpGs were located in CpG islands associated with

the genes *FGF12*, *OPLAH*, *COL25A1* and *METAP1D* (Table 2).

The panel developed on the GEO datasets was able to discriminate non-tumor from tumor samples with a sensitivity of 98.5% and a specificity of 98.8% at the 0.5 *B* value mean threshold (Fig. 3).

These sensitivities and specificities corresponded to an area under the ROC curve (AUC) of 0.998 (CI 0.994–1). These performances were reproducible with the same *B* value mean for the TCGA samples, with an AUC of 0.999 (CI 0.999–1). The difference in the AUC was not significant (two-sided Delong test, $p = 0.527$) (Fig. 4).

The specificity of the panel was compared to other tumor types in the TCGA with breast (non-tumor $n = 96$; tumor $n = 797$), lung (non-tumor $n = 74$; tumor $n = 346$), skin (non-tumor $n = 2$; tumor $n = 473$), stomach (non-tumor $n = 2$; tumor $n = 395$), esophagus (non-tumor $n = 16$; tumor $n = 184$), liver (non-tumor $n = 58$; hepatocarcinoma $n = 380$; cholangiocarcinoma $n = 34$) and pancreas (non-tumor $n = 10$; tumor $n = 185$) samples. The panel discriminated from other tumor types with an AUC of 0.912 (95% CI [0.896–0.927]), a sensitivity of 85.4% and a specificity of 84.8% (Fig. 5).

In the TCGA-COAD dataset, there were no significant differences in the methylation level of the panel according to age, ethnicity or sex (Supplementary data 2, 3 and 4).

Transcriptional analysis of the TCGA-COAD dataset revealed that *COL25A1* and *FGF12* were significantly downregulated between non-tumor and tumor samples. *OPLAH* did not show a significant change in transcription and *METAP1D* was significantly upregulated in tumor samples (Fig. 6).

There was a negative correlation between *COL25A1* and *FGF12* methylation and expression level, a positive correlation for *METAP1D* and no correlation for *OPLAH* (Supplementary data 5).

Tissues

qMSP primers were successfully designed and validated for three of the six potential biomarkers. The CpGs were cg07095995, cg08750504 and cg22882523, and CpGs were associated with *COL25A1*, *METAP1D* and *OPLAH*, respectively. The CpGs associated with these genes are indicated by the gene name. For the other three predicted biomarkers, no qMSP primers could be successfully designed due to the low region complexity, insufficient primer specificity, or secondary structures and the designed primers did not meet the required sensitivity or specificity for ddPCR (Table 3). The transposition of *COL25A1* methylated and *C1-Cless* primers to the Biorad QX200 EvaGreen Supermix[®] was successful, but *METAP1D* methylated primers lacked specificity and

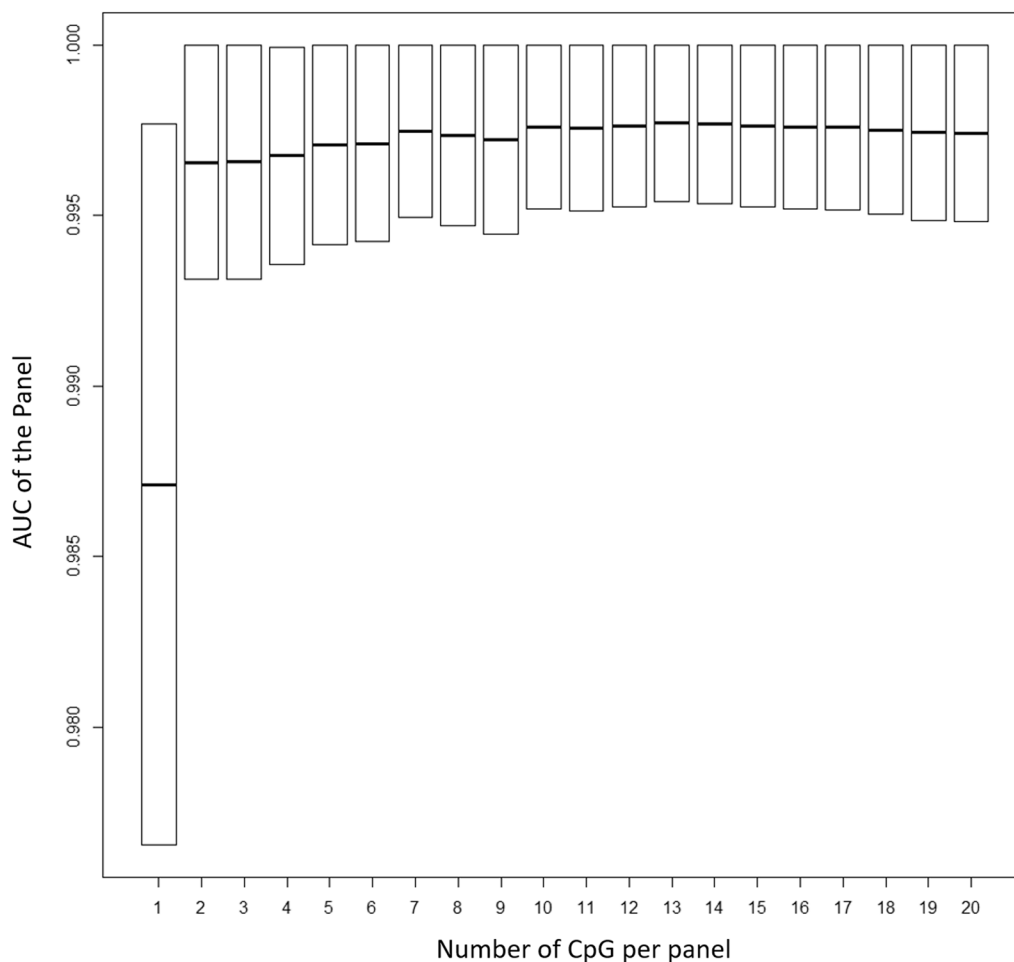


Fig. 2 Maximal AUC of the panel on the GEO datasets according to the number of CpGs in the panels. The performance of panels comprising up to twenty CpGs was assessed using the GEO datasets, with evaluation conducted based on the area under the curve (AUC). The AUC of the most effective panel is depicted by the black lines, while the 95% confidence intervals computed with Delong method are indicated by the boxes. The inclusion of a second CpG enhanced the panel’s performance, but additional CpGs did not yield further improvements

Table 2 Selected panel of 6 CpG with discriminating power for colorectal cancer

Probes’ name	Genomic location (hg19)	Associated gene	Location to gene
cg08002883	chr3:191227457	<i>FGF12</i>	Promotor
cg22882523	chr8:145107012	<i>OPLAH</i>	Intron 24/27
cg21535606	chr1:47911940	–	Intergenic*
cg07095995	chr4:110223980	<i>COL25A1</i>	Promotor
cg08750504	chr2:172946193	<i>METAP1D</i>	Exon 9/9
cg04921989	chr2:132183100	–	Intergenic*

*Location not associated with an annotation

had a high amount of primer dimers with this method. The addition of a probe containing an extra CpG to *METAP1D* increased the specificity of the ddPCR signal for the methylated DNA and eliminated the signal from the primer dimers. The *COL25A1* and *CI-Cless* primers had sufficient specificity for the methylated DNA and

primer dimers in the Biorad QX200 EvaGreen Supermix® conditions.

The methylation-specific primers of *COL25A1*, *METAP1D*, *OPLAH* and the DNA control *C-less* were tested by qMSP on 19 CRC and 15 non-tumor FFPE samples. The results confirmed a significantly higher

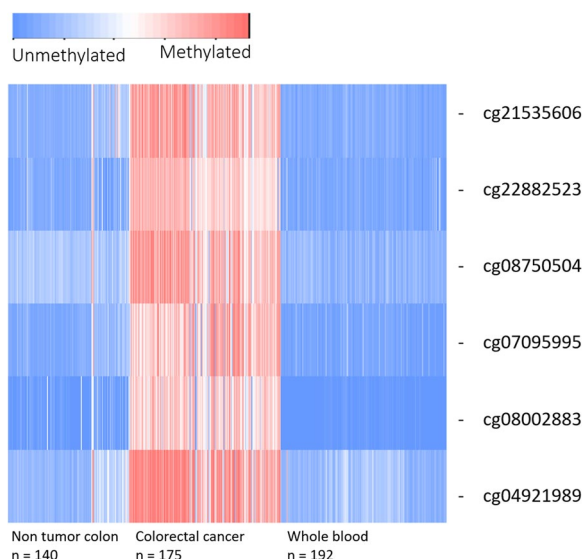


Fig. 3 Heatmap of the panel's *B* values on the GEO datasets. The panel's CpG differentiated non-tumor and whole blood from tumor samples (*n* = 140, 192 and 175, respectively). Non-tumor and blood samples were hypomethylated (blue) unlike the hypermethylated (red) tumor samples on the panel's CpG. However, no distinct sub-groups were identified though an unsupervised clustering process

methylation (Wilcoxon test, $p < 2.10^{-6}$ in each case) in the tumor samples compared to the paired non-tumor samples, with a combined sensitivity and specificity of 100% at a 0.5 *B* value threshold (Fig. 7).

Liquid biopsies

Of the three biomarkers validated in qMSP, only *COL25A1* and *METAP1D* could be transferred to the ddPCR system. The *OPLAH* qMSP primers did not generate a specific PCR amplicon under the ddPCR conditions. The AUCs for the combination of these two biomarkers on the GEO and TCGA datasets were 0.989 and 0.996, respectively (Supplementary data 6). Testing with the *OPLAH* primers did not produce signals under ddPCR conditions. Due to an insufficient amount of material, 4 CRC plasma samples were not tested for *COL25A1*, and 1 CRC plasma sample was not tested for *METAP1D*. *COL25A1* and *METAP1D* were both significantly more methylated in the plasma of CRC patients than in the plasma of healthy donors (Wilcoxon test, $p < 3.10^{-4}$ and $p = 0.024$, respectively). There were not significant effect of disease stage, age, gender and RAS/BRAF status on the plasma level of the biomarkers (Supplementary data 8–11).

COL25A1 and *METAP1D* both showed a specificity of 100% (95% CI [90–100], 0/35), and individual sensitivities of 35% (95% CI [20–54], 12/34) and 42% (95% CI [35–60], 13/31) at a threshold of 1 copy/μL, respectively. The thresholds were set in an elaboration cohort (*n* = 40) and confirmed in a validation cohort (*n* = 30). The combination of the mean methylated copy number of *COL25A1* and *METAP1D* showed an increase to 49% (95% CI [31–66], 18/35) sensitivity with a 100% (95% CI [90–100], 0/35) specificity and an AUC of 0.79 (95% CI [0.69–0.90] Delong test), as expected from the bioinformatics prediction (Fig. 8).

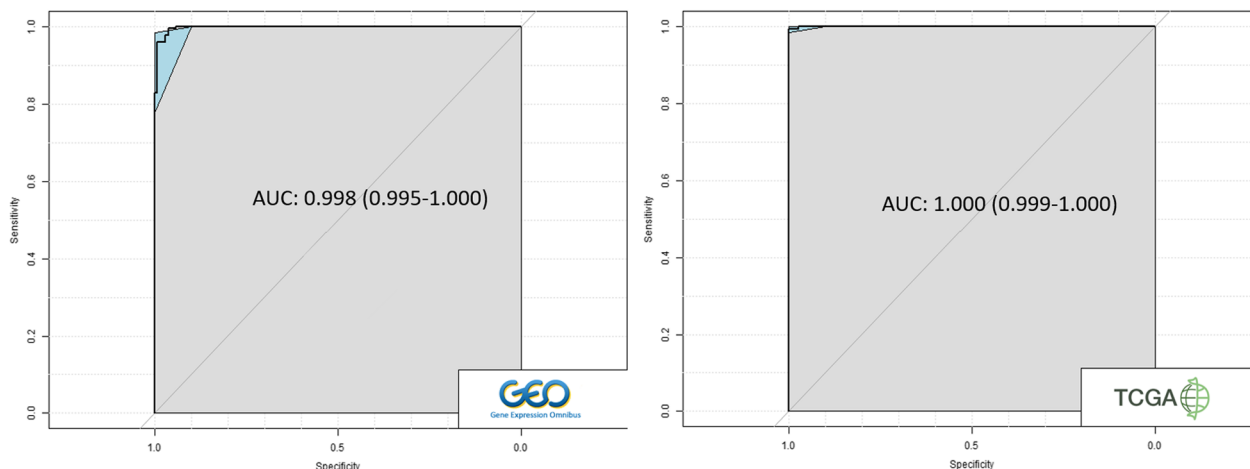


Fig. 4 ROC curves for the panel on both the GEO and TCGA datasets. ROC curves generated and the AUC calculated from the GEO and TCGA datasets. The ROC 95% confidence intervals computed with bootstrapping are depicted by the blue shapes. The panel's performance on the GEO dataset yielded an AUC of 0.998 and was confirmed on the TCGA dataset with an AUC of 1 and no significant statistical difference (bilateral Delong test, $p = 0.527$)

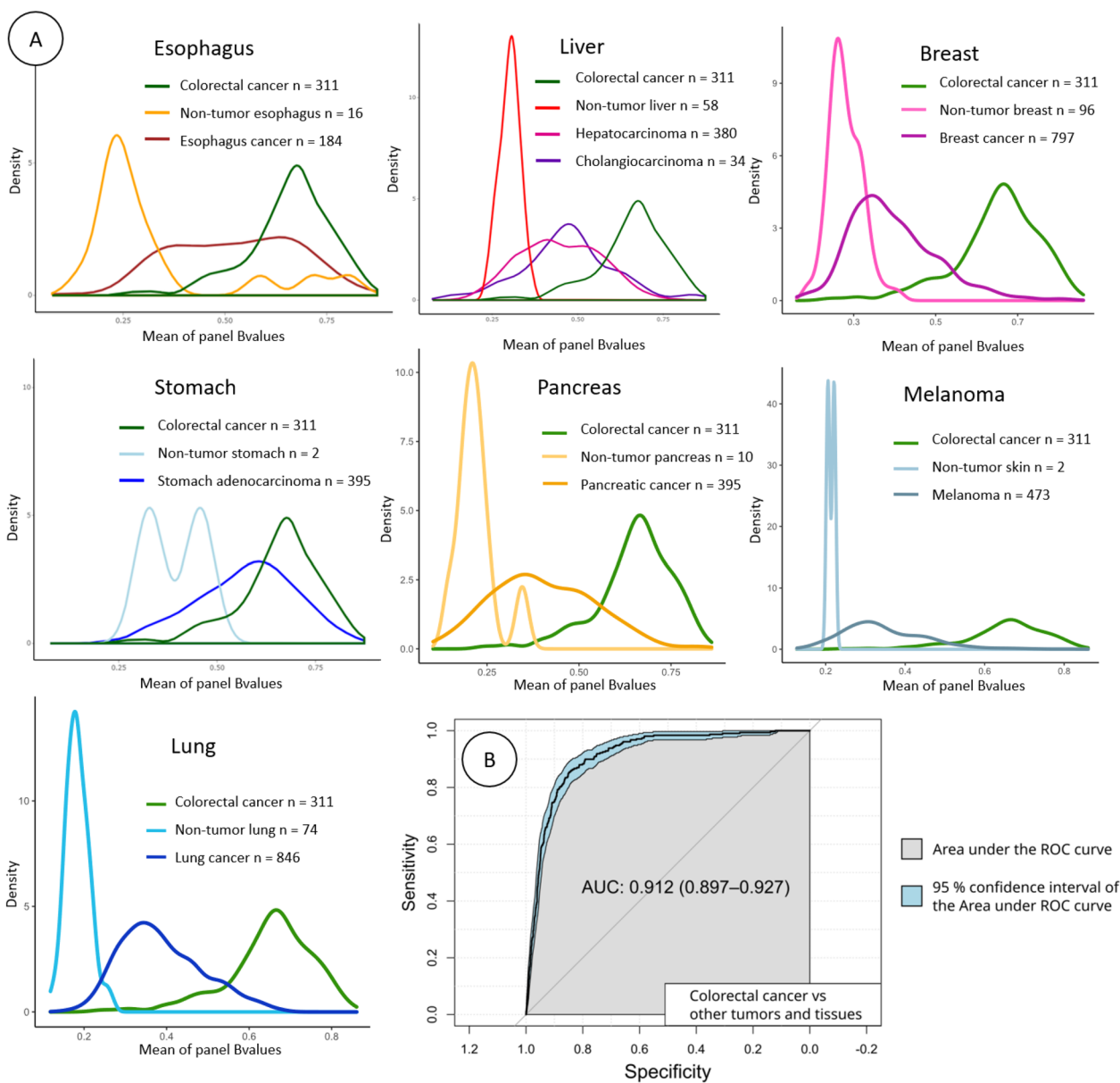


Fig. 5 Panel methylation profile in other tumor and non-tumor samples from the TCGA datasets. **A** The panel methylation means in colorectal cancer compared to other cancers from TCGA: pancreas, stomach, liver, esophagus, breast cancer, pancreatic cancer, lung cancer and melanoma. Additionally, colorectal cancer samples were compared to non-tumor tissues samples available in the TCGA datasets. In all comparisons, the colorectal cancer samples were more methylated than the other samples (with exception to non-tumor skin and non-tumor stomach samples due to the insufficient number of samples). **B** ROC curves and the AUC calculated from the colorectal samples compared to stomach, liver, esophagus, breast, lung, pancreas, melanoma and skin samples. ROC 95% confidence interval computed with bootstrapping are depicted by the blue shapes

Excluding the samples that were negative for the biomarkers and had fewer than 250 copies of C-Less per microliter, a higher proportion of CRC patient were detected (17/18) in the retained samples (Fig. 9).

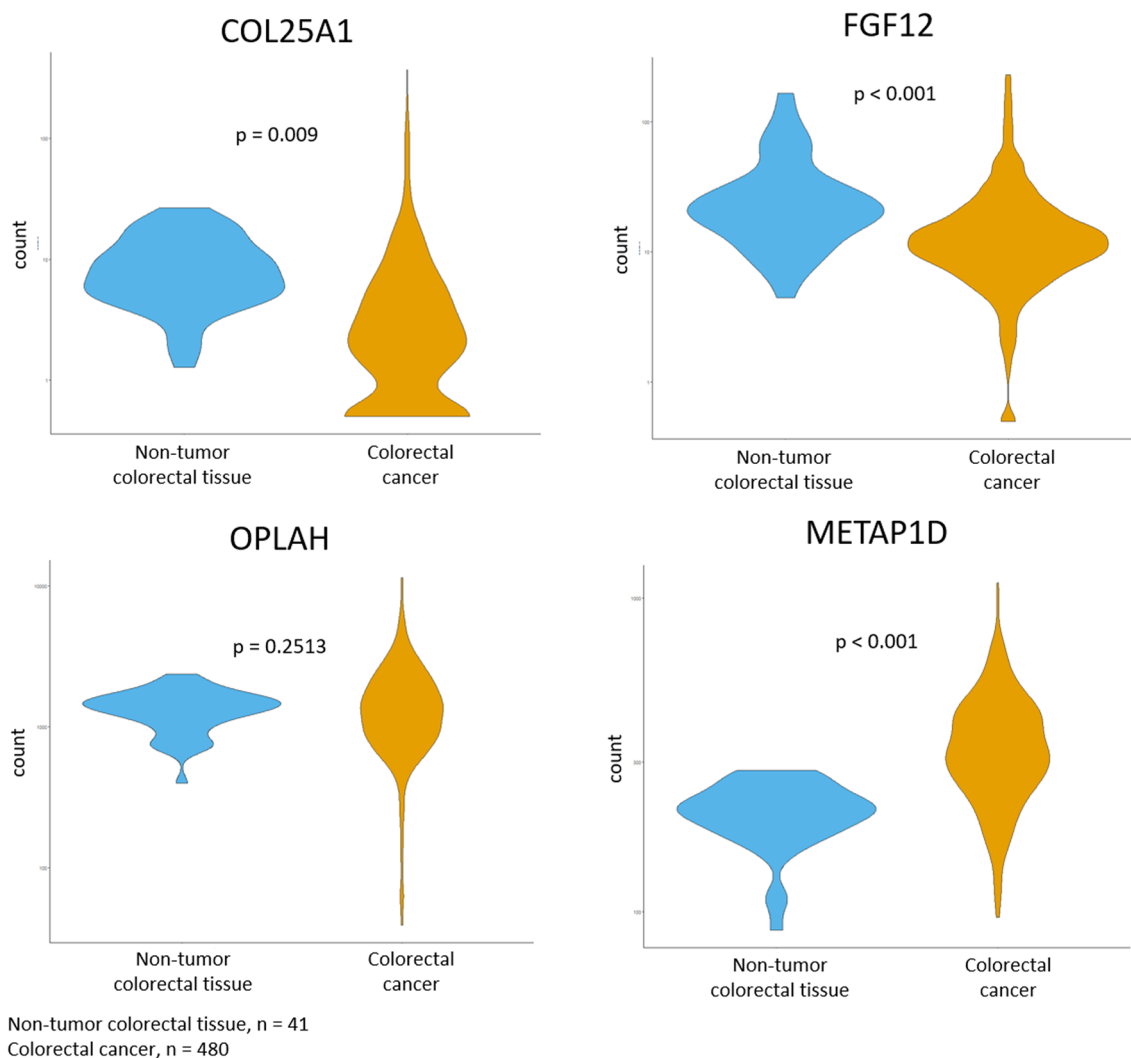


Fig. 6 Transcription profile of the panel related genes between tumor and non-tumor colorectal samples from the TCGA datasets. Violin plot of the transcription counts of the genes associated with the panel. Gene expression was measured in counts and was normalized using Deseq2. *COL25A1* and *FGF12* had significantly lower expression in colorectal cancer samples compared to non-tumor colorectal samples (Student test, $p = 0.009$ and < 0.001 , respectively). *OPLAH* did not demonstrate differential expression between the two sample types ($p = 0.2513$) and *METAP1D* had a significantly higher expression in colorectal cancer samples compared to non-tumor colorectal samples

Discussion

In our study, we show that *COL25A1* and *METAP1D* are CRC biomarkers that can be used in liquid biopsy.

In silico analysis

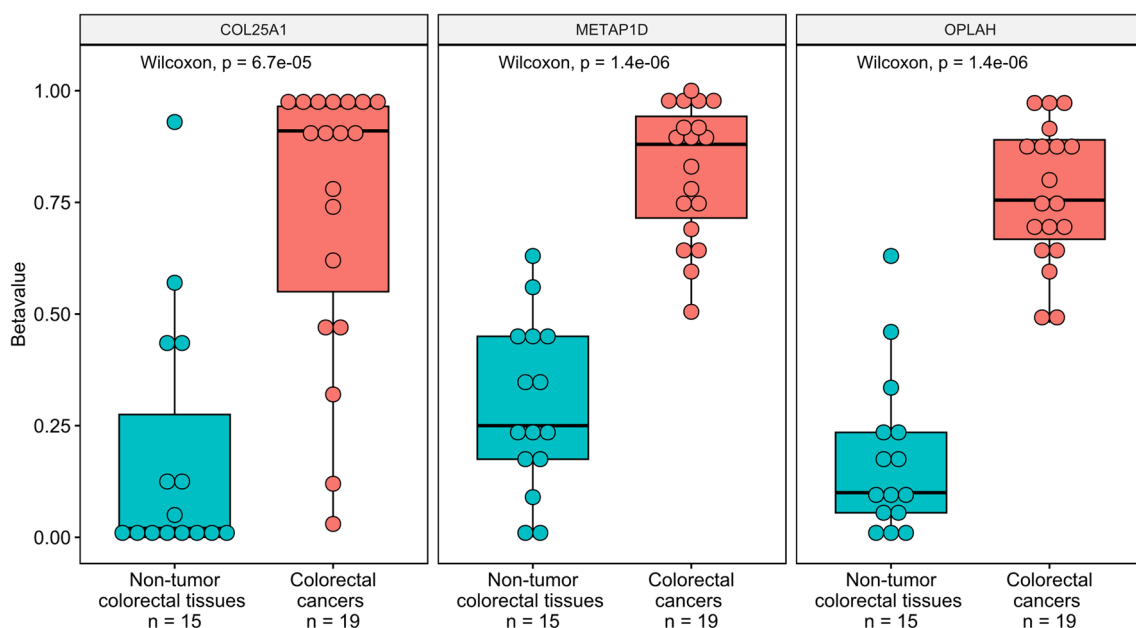
This work starts with a bioinformatics analysis of public databases. The originality of the bioinformatics analysis is that it searches for tumor and blood-specific methylation patterns specifically for use in liquid biopsy. This analysis was followed by the confirmation of the potential biomarkers in tissue samples by qMSP and then in plasma sample by ddPCR.

The panel is globally specific for the other cancer types and non-tumor tissues in the TCGA dataset. However, the panel is less discriminative for esophageal and gastric cancer with a specificity of 47.7%, as expected from the proximity in the human developmental lineage of colorectal, gastric, and esophageal tissues. These results suggest that the panel should be used with caution when used for diagnosis.

The technical limitations of transferring the markers from the Illumina Beadchip used in the databases to MSP and ddPCR must be considered during the bioinformatic analysis of the panel development. The Illumina Beadchip is based on sequence capture and single-base extension,

Table 3 Primers and probe of *COL25A1*, *METAP1D*, *OPLAH* and *C1-Cless*

MSP condition	Name	Forward	Reverse	Probe
Quantitative PCR	<i>COL25A1</i> <i>methylated</i>	AAT-TAA-TGT-TAT-TCG-CGA-AGT-CG	GAA-AAT-CCT-TCG-AAC-AAC-CG	–
	<i>Unmethylated</i>	GGA-ATT-AAT-GTT-ATT-TGT-GAA-GTT-GA	TAC-AAA-AAT-CCT-TCA-AAC-AAC-CAC	–
	<i>METAP1D</i> <i>methylated</i>	CGA-CCG-AAC-TAA-AAC-TCA-ACT-CG	GTT-TTA-GAG-AGA-GGA-GAT-ACG-GGC-G	–
	<i>Unmethylated</i>	TGT-ATG-TTG-TAG-GAG-TTG-TAA-ATG-T	CAA-CCA-AAC-TAA-AAC-TCA-ACT-CAC-A	–
	<i>OPLAH</i> <i>methylated</i>	TTT-TCG-TCG-TAG-TTC-GAA-GC	CTA-AAA-AAC-CGA-TAA-ACG-ACG-TA	–
	<i>Unmethylated</i>	GAG-TTT-TTT-GTT-GTA-GTT-TGA-AGT-G	AAA-TCC-TAA-AAA-ACC-AAT-AAA-CAA-CAT-A	–
Digital droplet PCR	<i>COL25A1</i> <i>methylated</i>	AAT-TAA-TGT-TAT-TCG-CGA-AGT-CG	GAA-AAT-CCT-TCG-AAC-AAC-CG	–
	<i>METAP1D</i> <i>methylated</i>	CGA-CCG-AAC-TAA-AAC-TCA-ACT-CG	GTT-TTA-GAG-AGA-GGA-GAT-ACG-GGC-G	6FAM-TGA-GAG-TAA-GGG-TTT-ATT-TGG-TCG-GGA-BHQ1
	Cless-C1	TTG-TAT-GTA-TGT-GAG-TGT-GGG-AGA-GAG-A	TTT-CTT-CCA-CCC-CTT-CTC-TTC-C	–

**Fig. 7** Quantitative methylation specific PCR of *COL25A1*, *METAP1D* and *OPLAH* on tumor and non-tumor colorectal tissues samples. The methylation level of *COL25A1*, *METAP1D* and *OPLAH* was significantly higher in tumor compared to non-tumor tissues samples (Wilcoxon test, $p < 10^{-4}$ for each biomarker)

whereas MSP is based on differential hybridization of primers. The methylation-specific hybridization temperature can be challenging because too low temperature can induce non-specific hybridization, and too high temperature can prevent the hybridization. In fact, four of the six biomarkers could not be validated in plasma with MSP due to limitations of the technique transposition. Considering the loss of targets due to technical feasibility,

the in silico panel size was set at six CpGs resulting in two biomarkers (*COL25A1* and *METAP1D*) that could have been transposed in ddPCR.

Biological validation

The biological validation of these biomarkers was performed using a cohort of tissue and plasma samples. A limitation of this study is that the included patients

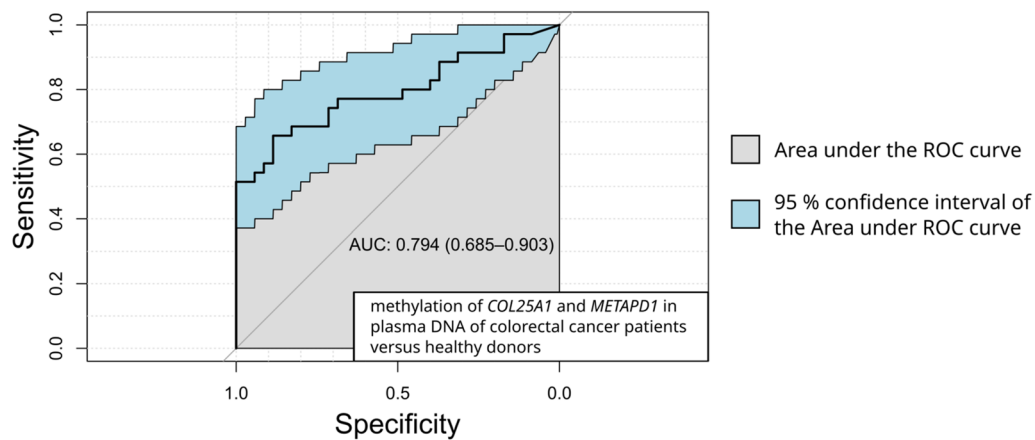


Fig. 8 ROC curves of the combination of *COL25A1* and *METAP1D* on plasma samples. ROC's curves of the mean of the copy number of methylated *COL25A1* and *METAP1D* on plasma samples. Plasma samples were collected from colorectal cancer patients ($n=35$) and healthy donors ($n=35$). The ROC 95% confidence interval computed with bootstrapping is shown by the blue shape

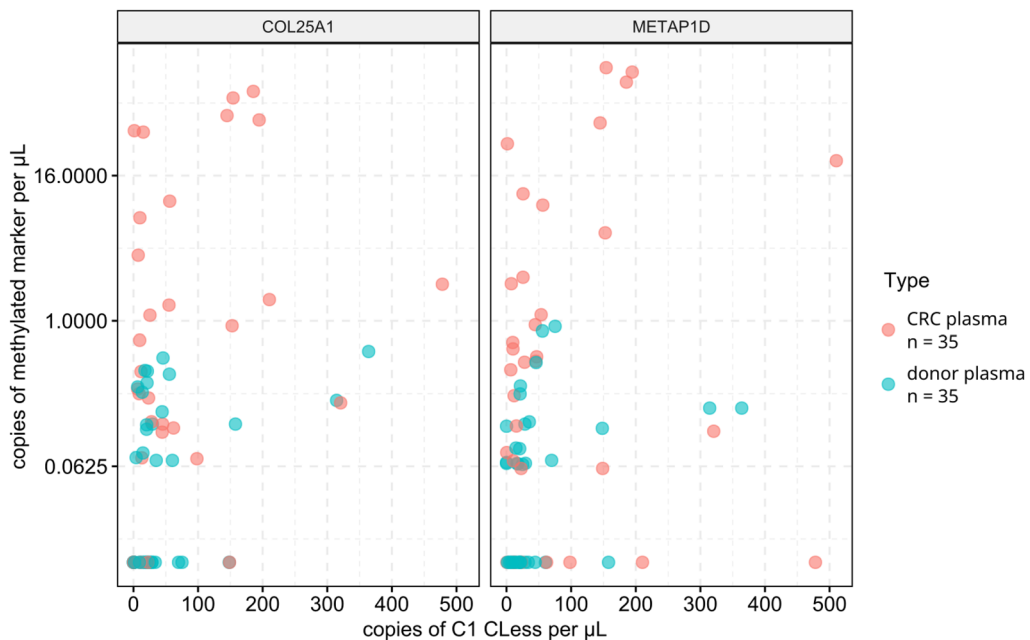


Fig. 9 *COL25A1* and *METAP1D* methylation (copy/µL) by cellular control in plasma samples. The number of methylated copies were significantly higher in the plasma samples of patients with metastatic colorectal cancer ($n=35$) compared to plasma samples of healthy donor ($n=35$) for both *COL25A1* and *METAP1D* (Wilcoxon test, $p < 3.10^{-4}$ and 0.024, respectively). Specificity for both biomarkers was 100%, the sensitivity was 35% for *COL25A1*, 42% for *METAP1D* and 49% for the combination of *COL25A1* and *METAP1D*

were all with advanced CRC stage III or IV. A future work would be needed to investigate the sensitivity and specificity for earlier CRC stages. However, there were no significant differences in the plasma levels of the biomarkers between stages III and IV (Supplementary Data 8). Another limitation of this study is that the healthy donors are not fully representative of the population in which the biomarkers would be used

for follow-up. However, this study serves as a proof of concept for biomarker discovery, where the biomarkers were evaluated in both healthy individuals and advanced CRC cases to determine their inherent sensitivity and specificity. Future prospective studies are needed to validate these biomarkers in relevant clinical settings, such as patients with colorectal dysplasia, other neoplasms and non-tumor diseases.

The main experimental limitation was the design of MSP primers in CpG islands due to the low complexity region and high CpG content, which increase the melting temperature and limit the primer size and specificity. The primers also had to work with the Bio-Rad's ddPCR conditions, which are only partially flexible due to the droplet generation by emulsion.

Panel scoring differs in the plasma ddPCR assay compared to the tissue qMSP and the bioinformatic analysis. This difference is due to the typically highly diluted ctDNA within the non-tumor cell-free DNA in the plasma samples. The expected B values would all have been extremely low regardless of the presence or absence of ctDNA. Since ddPCR primarily provides the absolute methylated copy number and the difference is expected to be more important, the copy number was chosen over the B value for the panel scoring in the plasma.

The positive cut-off for the biomarkers was determined from the ROC curves and rounded to 1 copy per microliter. The positive cut-off for the Cless control was set at 250 copies per microliter, which means that at the Cless limit, the minimal tumor DNA fraction is 0.4%, which is already below around the limit of detection of most NGS technologies [24]. These cutoffs were determined from the experimental data, and they should be confirmed in larger cohorts in future studies.

A CRC plasma sample is negative for the biomarkers but positive for the Cless, which is a false negative. This result could be due to the lack of sensitivity of the biomarkers or may be due to the absence of CRC DNA in the plasma sample. Such results are expected with the low levels of circulating tumor DNA.

Molecular characteristics of the tissue samples were not available; *KRAS*, *NRAS*, and *BRAF* mutation status did not significantly affect biomarker methylation levels in the plasma cohort (Supplementary data 10). The GSE48684 GEO series provided a comprehensive molecular characterization of the CRC in their publication. They included 64 CRC samples, of which 9 had MSI status, 29 *KRAS* and 9 *BRAF* with hotspot mutations. All these samples were positive for *COL25A1* and *METAP1D* methylation, suggesting that these biomarkers may be independent of the molecular characteristics of CRC. Another limitation of this study is that the biomarkers were not tested on the rare histologic subtypes of CRC that represent approximately 10% of CRC, such as mucinous adenocarcinoma, signet-ring cell carcinoma, and medullary carcinoma [25].

Biological role of CRC epigenetic biomarkers

In support of our analysis, *FGF12* and *OPLAH* have been shown to be hypermethylated in CRC [26, 27]. The biological function of *FGF12* has not been clearly

established, but murine models suggest that *FGF12* plays a protective role against radiation-induced lesions in the intestine [28]. *OPLAH* encodes 5-oxoprolinase, an enzyme involved in the gamma-glutamyl cycle of glutathione metabolism. The transcription of *OPLAH* is not correlated with the CpG methylation of the panel and is explained by the location of the CpG in the gene far from the promoter, intron 24 or exon 9 depending on the considered transcript. However, the promoter of *EXOCS4*, a potential colorectal oncogene [29], is located 20 kb from the *OPLAH* CpG. *EXOCS4* transcription was increased in tumor samples (two-sided Student's t test, $p < 0.001$). CpG methylation may be involved in the transcription of this potential oncogene.

The *COL25A1* methylation has been associated with the transformation of cervical intraepithelial neoplasia [30]. The transcription initiation site of *COL25A1* is associated with a divergent transcript, *COL25A1-DT*. This divergent transcript has been associated with astrocytoma in the MalaCards database [31], but *COL25A1-DT* transcription was not significantly different between CRC and non-tumor tissues in the TCGA-COAD dataset (two-sided Student's t test, $p = 0.78$), suggesting that this divergent transcript does not play a role in CRC.

METAP1D is one of the two isoforms of methionine aminopeptidase that catalyzes the excision of N-terminal methionine from nearly 70% of newly synthesized proteins for protein stability, cellular location, and activation [32]. *METAP1D* plays a role in the cell cycle G2/M phase [33] and *METAP1D* has been described to be overexpressed in CRC, lung and breast cancer [34] and cellular models suggest that it is a potential oncogene [35]. The functional and pathological aspects of the genes associated with the CpG panel suggest a biological relevance of the methylation pattern.

Comparison of the biomarkers with other DNA methylation CRC biomarkers

Two plasma-based CRC DNA methylation tests are FDA-approved for CRC screening, the detection of the *SEPT9* methylation (epi Procolon[®]) [36] and a combined test including DNA methylation (Shield[®]) [11]. *SEPT9* methylation has a sensitivity of 46.0% for stage III and 77.4% for stage IV and a specificity of 91.5%, which is similar to the performance of the *COL25A1* and *METAP1D*. Shield[®] has a sensitivity of 83.1% and a specificity of 89.9% for all stages, which is higher than *COL25A1* and *METAP1D*, but Shield[®] relies on multiple techniques, and may be less applicable than a ddPCR-based technique.

The performance of the biomarkers is similar to other published plasma-based DNA methylation biomarkers using ddPCR such as the combination of *SDC2* and *NPY*

(sensitivity of 33–54%, specificity of 72–96%); *IKZF1* and *SEPT9* (sensitivity of 19–42%, specificity of 88–96%) [37]; *LIFR* and *ZNF304* (sensitivity of 70%, specificity of 92%) [38]; *C9orf50*, *KCNQ5* and *CLIP4* (sensitivity of 85% and specificity of 99%) [39]. However, *COL25A1* and *METAP1D* show 100% specificity in the plasma cohort; this excellent specificity may provide a better positive predictive value and therefore a better clinical utility, which needs to be validated in future studies.

Clinical uses of CRC epigenetic biomarkers

In this work, based on the analysis of two databases, a panel of biomarkers CRC designed for liquid biopsy was identified by bioinformatics analysis and subsequently confirmed by experiments on healthy donors and patients with CRC.

These results are compatible with the clinical applications of biomarkers, including the follow-up of CRC patients. In follow-up, ctDNA can be used to monitor minimal residual disease after surgery. Methylation biomarkers are particularly useful for the follow-up of patients, regardless of the *RAS*, *BRAF* and MSI status. The use of these biomarkers could potentially be implemented in diagnosis of CRC patients for whom by tissue biopsy is not possible due to clinical limitations such as hemorrhage or general anesthesia risks. In these cases, ctDNA can be useful to detect cancer-associated and actionable variants. However, if no variants are found, this may be due to a lack of variants of interest in the tumor or insufficient ctDNA in the sample. Highly specific methylation biomarkers can reduce the number of resamples and the time to decision-making. However, future studies are needed to evaluate the utility of the markers in diagnostic and screening settings.

In in the TCGA dataset, the *in silico* performance of *COL25A1* and *METAP1D* is comparable to *SEPT9* methylation (Supplementary data 7), which is FDA-approved for CRC screening, indicating that these biomarkers could be used for blood-based screening. DNA methylation alteration occurs early in the CRC oncogenesis, which is a strong argument for its usefulness in CRC screening. However, a limitation of blood-based biomarkers for screening and follow-up is that the amount of ctDNA is likely to correlate with CRC tumor size [40]. This correlation may affect the sensitivity of blood-based cancer biomarkers. Smaller CRC tumors may have reduced sensitivity [41], which reduces the negative predictive value. Achieving high specificity in blood-based biomarkers is critical to ensure clinically relevant positive predictive values. Another aspect is that biomarkers do not require fecal manipulation. We believe that a blood-based biomarker would have a high level of acceptance in the target population. Since

CRC screening is typically repeated, the low sensitivity resulting from the low concentration of ctDNA could be mitigated through repeated screening over time.

In this study, we performed a bioinformatics analysis of methylation databases for biomarker discovery and validated them in liquid biopsies. Future studies are needed to evaluate the application of these biomarkers in routine medical practice. Further investigations may lead to the discovery of novel biomarkers for other cancers types or even a pan-cancer panel. However, multicenter studies with larger cohorts could provide a more accurate assessment of the clinical performance of the panel.

Conclusion

This study discovered and validated two new CRC biomarkers that can be used in liquid biopsy using ddPCR. These biomarkers have promising performance metrics. However, further studies are needed to evaluate the clinical application of these biomarkers. Nevertheless, such analyses can be repeated for other cancer types, and it may be challenging to discover a DNA methylation panel that covers all or most cancer types.

Abbreviations

CRC	Colorectal cancer
ctDNA	Circulating tumor DNA
ddPCR	Digital droplet PCR
GEO	Gene Expression Omnibus
TCGA	The Cancer Genome Atlas
AUC	Area under the ROC curve
qMSP	Quantitative methylation-specific PCR
FFPE	Formalin-fixed, paraffin-embedded
FDA	Food and Drug Administration

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-024-01748-1>.

Additional file 1

Acknowledgements

We thank all the patients and donors who agreed to participate in this study. We thank the EPIgenetics and GENE EXPRESSION Technical Platform (EPIGEN-Exp) for making the Bio-Rad digital PCR device available, their expertise for the design of the experiments, and their help for the results analysis. This study was also supported by CELIA grant of Besançon university hospital.

Author contributions

AO, ZS, and JPF contributed to the conception and design. AO, EH, PP, ZS, and JPF contributed to the development of methodology. AO, CM, FM, MH, CB, AV, JV, and ZS acquired the data. AO, JD, JV, ZS, JPF, PP, EH, and MG analyzed and interpreted the data. AO, PP, EH, CM, FM, JD, AV, JV, MH, CB, JPF, ZS, and MG participated in writing, review and/or revision of the manuscript. AO, ZS, and JPF contributed to the study supervision. All authors read and approved the final manuscript.

Availability of data and materials

The datasets generated and analyzed during this study (birthdate, admission date, discharge date, date of death, etc.) are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the National Research Committee and with the 1964 Helsinki Declaration and its later amendments. In France, this search is considered as a non-interventional study according to European legislation. All patients were individually informed that their data should be used for scientific research. All the experimental protocols were approved by the scientific board of the regional biobank of Franche-Comté, France (registration number BB-0033-00024, Tumorothèque Régionale de Franche-Comté) that ensures patients informed consent.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Oncobiology, University Hospital of Besançon, 3 Boulevard Alexandre Fleming, 25000 Besançon, France. ²UMR1098, INSERM, University of Bourgogne Franche-Comté, Besançon, France. ³Department of Pathology, University Hospital of Besançon, 25000 Besançon, France. ⁴Department of Oncology, University Hospital of Besançon, 25000 Besançon, France. ⁵Laboratory of Experimental Pathology, GIGA-Cancer, University of Liege, Liège, Belgium.

Received: 8 July 2024 Accepted: 16 September 2024

Published online: 18 October 2024

References

- Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. Global cancer observatory: cancer today. Lyon: International Agency for Research on Cancer; 2024. Available from: <https://gco.iarc.who.int/today>, Accessed 06 March 2024.
- Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet*. 2019;394:1467–80.
- Winawer SJ, Zauber AG. The advanced adenoma as the primary target of screening. *Gastrointest Endosc Clin N Am*. 2002;12(1–9):v.
- Tinmouth J, Lansdorp-Vogelaar I, Allison JE. Faecal immunochemical tests versus guaiac faecal occult blood tests: what clinicians and colorectal cancer screening programme organisers need to know. *Gut*. 2015;64:1327–37.
- Malla M, Loree JM, Kasi PM, Parikh AR. Using circulating tumor DNA in colorectal cancer: current and evolving practices. *J Clin Oncol*. 2022;40:2846–57.
- Morris VK, Kennedy EB, Baxter NN, Benson AB, Cercek A, Cho M, et al. Treatment of metastatic colorectal cancer: ASCO guideline. *J Clin Oncol*. 2023;41:678–700.
- Levin-Sparenberg E, Bylsma LC, Lowe K, Sangare L, Fryzek JP, Alexander DD. A systematic literature review and meta-analysis describing the prevalence of KRAS, NRAS, and BRAF gene mutations in metastatic colorectal cancer. *Gastroenterol Res*. 2020;13:184–98.
- You JS, Jones PA. Cancer genetics and epigenetics: Two sides of the same coin? *Cancer Cell*. 2012;22:9–20.
- Weinstein JN, Collisson EA, Mills GB, Shaw KM, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet*. 2013;45:1113–20.
- Müller D, Györfy B. DNA methylation-based diagnostic, prognostic, and predictive biomarkers in colorectal cancer. *Biochim Biophys Acta Rev Cancer*. 2022;1877:188722.
- Chung DC, Gray DM, Singh H, Issaka RB, Raymond VM, Eagle C, et al. A cell-free DNA blood-based test for colorectal cancer screening. *N Engl J Med*. 2024;390:973–83.
- Diaz LA, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol*. 2014;32:579–86.
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. [Internet]. [cited 2021 Jan 22]. Available from: <https://www.R-project.org/>
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform*. 2011;12:77.
- Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 2007;23:1846–7.
- Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016;44:e71–e71.
- Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinform*. 2010;11:587.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–45.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
- Stevenson M, Sergeant E, Heuer C, Nunes T, Heuer C, Marshall J, et al. epiR: Tools for the Analysis of Epidemiological Data [Internet]. 2024 [cited 2024 Aug 25]. Available from: <https://cran.r-project.org/web/packages/epiR/index.html>
- Li L-C, Dahiya R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics*. 2002;18:1427–31.
- Tusnády GE, Simon I, Váradi A, Arányi T. BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Res*. 2005;33:e9.
- Weisenberger DJ, Trinh BN, Campan M, Sharma S, Long TI, Ananthnarayan S, et al. DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. *Nucleic Acids Res*. 2008;36:4689–98.
- Singh RR. Next-generation sequencing in high-sensitive detection of mutations in tumors: challenges, advances, and applications. *J Mol Diagn*. 2020;22:994–1007.
- BlueBooksOnline [Internet]. [cited 2024 Sep 5]. Available from: <https://tumourclassification.iarc.who.int/chaptercontent/31/280>.
- Li H, Du Y, Zhang D, Wang L-N, Yang C, Liu B, et al. Identification of novel DNA methylation markers in colorectal cancer using MIRA-based microarrays. *Oncol Rep*. 2012;28:99–104.
- Naumov VA, Generozov EV, Zaharjevskaya NB, Matushkina DS, Larin AK, Chernyshov SV, et al. Genome-scale analysis of DNA methylation in colorectal cancer using Infinium HumanMethylation450 BeadChips. *Epigenetics*. 2013;8:921–34.
- Nakayama F, Umeda S, Yasuda T, Fujita M, Asada M, Meineke V, et al. Cellular internalization of fibroblast growth factor-12 exerts radioprotective effects on intestinal radiation damage independently of FGFR signaling. *Int J Radiat Oncol Biol Phys*. 2014;88:377–84.
- Pan Y, Tong JHM, Kang W, Lung RWM, Chak WP, Chung LY, et al. EXOSC4 functions as a potential oncogene in development and progression of colorectal cancer. *Mol Carcinog*. 2018;57:1780–91.
- Lendvai A, Johannes F, Grimm C, Eijsink JHH, Wardenaar R, Volders HH, et al. Genome-wide methylation profiling identifies hypermethylated biomarkers in high-grade cervical intraepithelial neoplasia. *Epigenetics*. 2012;7:1268–78.
- Rappaport N, Fishilevich S, Nudel R, Twik M, Belinky F, Plaschkes I, et al. Rational confederation of genes and diseases: NGS interpretation via GeneCards. *MalaCards VarElect Biomed Eng Online*. 2017;16:72.
- Lee Y, Kim H, Lee E, Hahn H, Heo Y, Jang DM, et al. Structural insights into N-terminal methionine cleavage by the human mitochondrial methionine aminopeptidase, MetAP1D. *Sci Rep*. 2023;13:22326.
- Hu X, Addlagatta A, Lu J, Matthews BW, Liu JO. Elucidation of the function of type 1 human methionine aminopeptidase during cell cycle progression. *Proc Natl Acad Sci*. 2006;103:18148–53.
- Randhawa H, Chikara S, Gehring D, Yildirim T, Menon J, Reindl KM. Overexpression of peptide deformylase in breast, colon, and lung cancers. *BMC Cancer*. 2013;13:321.
- Leszczyniecka M, Bhatia U, Cueto M, Nirmala NR, Towbin H, Vattay A, et al. MAP1D, a novel methionine aminopeptidase family member is overexpressed in colon cancer. *Oncogene*. 2006;25:3471–8.
- Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut*. 2014;63:317–25.

37. Petit J, Carroll G, Williams H, Pockney P, Scott RJ. Evaluation of a multi-gene methylation blood-test for the detection of colorectal cancer. *Med Sci (Basel)*. 2023;11:60.
38. Li D, Zhang L, Fu J, Huang H, Liu Y, Zhu L, et al. Discovery and validation of tissue-specific DNA methylation as noninvasive diagnostic markers for colorectal cancer. *Clin Epigenet*. 2022;14:102.
39. Jensen SØ, Øgaard N, Ørntoft M-BW, Rasmussen MH, Bramsen JB, Kristensen H, et al. Novel DNA methylation biomarkers show high sensitivity and specificity for blood-based detection of colorectal cancer—a clinical biomarker discovery and validation study. *Clin Epigenet*. 2019;11:158.
40. Strijker M, Soer EC, de Pastena M, Creemers A, Balduzzi A, Beagan JJ, et al. Circulating tumor DNA quantity is related to tumor volume and both predict survival in metastatic pancreatic ductal adenocarcinoma. *Int J Cancer*. 2020;146:1445–56.
41. Avanzini S, Kurtz DM, Chabon JJ, Moding EJ, Hori SS, Gambhir SS, et al. A mathematical model of ctDNA shedding predicts tumor detection size. *Sci Adv*. 2020;6:4308.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.