Kuerban *et al. World Journal of Surgical Oncology (2025) 23:90*
https://doi.org/10.1186/s12957-025-03747-9

World Journal of Surgical Oncology

RESEARCH

Open Access

cfDNA hydroxymethylcytosine profiling for detection metastasis and recurrence of Esophageal Squamous Cell Carcinoma

Subinuer Kuerban[1†], Hangyu Chen[2,3†], Long Chen[2,3†], Lei Zhang[2,3], Xuehui Li[1], Baixin Zhen[1], Hong Xiao[4], Yingzhu Chen[7], Haitao Zhou[8], Zhen Liang[8], Guobing Xu[7], Yicun Tao[1*], Jian Lin[2,3,4,5*] and Xiaozheng Kang[6*]

Abstract

Background A blood-based approach to monitor metastasis and recurrence of esophageal squamous cell carcinoma (ESCC) remains undeveloped. This study aimed to establish a dependable model utilizing cfDNA 5-hydroxymethylcytosines (5hmC) signatures to detect these conditions in ESCC.

Methods The 5hmC-Seal technique was employed to generate comprehensive 5hmC profiles from the plasma cell-free DNA (cfDNA) of 122 ESCC patients, classified into 72 with metastasis, 50 without metastasis, 30 with recurrence, and 92 without recurrence. Initial steps involved identifying distinct hydroxymethylation signatures linked to metastasis and recurrence. Machine learning algorithms were then utilized to construct predictive models.

Results The study confirmed that 5hmC-based markers are predictive of metastasis and recurrence among ESCC patients. The analysis of 14 5hmC biomarkers revealed a sensitivity of 88.90% and a specificity of 84.00% (AUC = 0.922) in differentiating patients with ESCC metastasis from those without in the validation cohort. Similarly, 11 5hmC biomarkers showed a sensitivity of 93.30% and a specificity of 89.10% (AUC = 0.936) in identifying recurrent versus non-recurrent ESCC cases. Additionally, a wp-score for metastasis and recurrence, derived from the 5hmC marker, prognosticated patient outcomes.

Conclusions The findings indicate that 5hmC markers from cfDNA serve as effective epigenetic indicators for the non-invasive detection of ESCC metastasis and recurrence.

Keywords 5-hydroxymethylcytosine, Esophageal squamous cell carcinoma, Liquid biopsy, Metastasis, Recurrence

†Subinuer Kuerban, Hangyu Chen and Long Chen contributed equally to this work.

*Correspondence:
Yicun Tao
taoyicun@xjmu.edu.cn
Jian Lin
linjian@pku.edu.cn
Xiaozheng Kang
kangxz@cicams.ac.cn
[1] School of Pharmacy, Xinjiang Medical University, Urumqi 830017, China
[2] Department of Pharmacy, Peking University Third Hospital, Beijing 100191, China
[3] Peking University, Third Hospital Cancer Center, Beijing 100191, China
[4] Key Laboratory of Tropical Biological Resources of Ministry of Education, School of Pharmaceutical Sciences, Hainan University, Haikou 570100, China
[5] Synthetic and Functional Biomolecules Center, Peking University, Beijing, China
[6] Section of Esophageal and Mediastinal Oncology, Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[7] Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Clinical Laboratory, Peking University Cancer Hospital & Institute, Beijing, China
[8] Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), the First Department of Thoracic Surgery, Peking University Cancer Hospital and Institute, Peking University School of Oncology, Beijing, China

Introduction

Esophageal cancer ranks among the top cancers globally, with China contributing to nearly half of the worldwide occurrences, including new cases and fatalities annually. Predominantly, these cases involve esophageal squamous cell carcinoma (ESCC), which constitutes over 90% of such instances [1]. The primary contributors to the grim prognosis of ESCC are metastasis and the recurrence after treatment [2]. There is a critical demand for innovative technologies capable of detecting ESCC metastasis and recurrence.

While conventional diagnostics like serum tumor markers, imaging, endoscopy, and histopathological evaluations are prevalent [3, 4], their efficacy, particularly that of carcinoembryonic antigen (CEA) and carbohydrate antigen 19–9 (CA19-9), along with ultrasound and CT imaging, is compromised by low sensitivity and specificity in detecting metastasis and recurrence [5–7]. This underscores the imperative for novel molecular markers to track these conditions in ESCC.

Liquid biopsy, which involves the sampling and analysis of non-solid biological tissues (e.g., urine, saliva, ascites, or cerebrospinal fluid), provides a swift, non-invasive alternative to tissue biopsies, beneficial for ongoing tumor assessments, diagnosis, and monitoring [8–10]. Innovations in cell-free DNA (cfDNA) have shown potential to transform cancer screening, diagnosis, and treatment, facilitating a 'liquid biopsy' that allows for the molecular examination of solid tumors. This technique, being non-invasive, enables continuous and dynamic monitoring of tumor molecular dynamics, presenting clear advantages over traditional tissue biopsies.

Within the genomic landscape, 5hmC represents a significant class of epigenetic modifications that arise from the oxidation of 5-methylcytosines through the action of ten-eleven translocation enzymes [11–14]. The alterations observed in promoters, gene bodies, and regulatory elements, such as enhancers, reflect the activation of genes within mammalian genomes and are particularly effective for identifying the activation of specific genes or loci [12]. Recent studies have demonstrated that alterations in 5hmC are crucial to the pathobiology of cancer, particularly marked by a significant decrease in 5hmC levels across a range of malignancies [15–23]. Consequently, 5hmC has been recognized as an emerging class of epigenetic markers in cancer, showing significant potential for application in precision medicine due to its relevance to cancer genetics, tissue-specific expression, and crucially, the technological advancements that enable its utilization in non-invasive liquid biopsies [24–28].

In the current research, by applying the precise and dependable 5hmC-Seal approach, we charted the genome-wide patterns of cfDNA 5hmC profiles among 122 ESCC patients, establishing a predictive model for the identification of metastasis and recurrence. The efficacy of this model was confirmed through assessments of its sensitivity and specificity in recognizing metastasis and recurrence. Moreover, we explored the relationship between the predictive scores, calculated from the cfDNA 5hmC model, and various clinical parameters. The outcomes suggest that 5hmC markers in cfDNA are capable of effectively tracking metastasis and recurrence in ESCC.

Methods

Sample size calculation

Sample size calculations were based on guidelines from Hajian-Tilaki [29]. The necessary sample size is determined by the following formula:

$$N = \frac{Z^2_{\frac{\alpha}{2}} V\left(\widehat{AUC}\right)}{d^2}$$

Here, α is computed in the following manner, and ϕ^{-1} represents the inverse of the standard cumulative normal distribution, given a pre-established AUC of 0.9:

$$\alpha = \varphi^{-1}(0.9) \times 1.414 = 1.281552 \times 1.414 = 1.812115$$

the $V\left(\widehat{AUC}\right)$ can be driven as follows:

$$V\left(\widehat{AUC}\right) = \left(0.0099 \times e^{\frac{-\alpha^2}{2}}\right) \times \left(6\alpha^2 + 16\right) = 0.068435$$

To calculate the AUC with 95% certainty and a precision margin of approximately 0.05, the requisite sample size is derived by inserting $V\left(\widehat{AUC}\right)$ values and adjusting d to 0.05:

$$N = \frac{\left(1.96^2 \times 0.068435\right)}{0.05^2} = 105$$

which means there were 105 samples needed for this study.

Participants and study design

This investigation included 122 ESCC patients from the Peking University Cancer Hospital, enrolled from 2018 to 2020. Each participant consented to the study. The inclusion criteria were: individuals aged over 18, clinically confirmed as having ESCC via endoscopy, pathological biopsy, and requisite imaging tests according to the 8th edition of the TNM classification by the American Joint Committee on Cancer/Union for International Cancer Control (AJCC/UICC). All subjects had comprehensive clinical records and provided essential information for the research. Exclusion criteria encompassed: severe cardiac, liver, or kidney diseases; other concurrent cancers; coagulopathy; significant depression or other psychological disorders; chronic ailments such as diabetes and hypertension; conditions of pregnancy or lactation; autoimmune disorders; severe infections. Definitions for metastasis were categorized as follows: Whereas N0 signified the absence of metastasis to any regional lymph nodes (RLN), N1 indicated the presence of metastasis to 1–2 RLN, N2 signified the presence of metastases to 3–6 RLN, and N3 involved seven or more RLN. Definition for recurrence: recurrence diagnosis was made according to the surgical records and postoperative regular review of imaging and pathological data. The specific criteria are as follows: puncture pathology clear lymph node metastasis; CT showed lymph node short diameter ≥ 10 mm or paraesophageal, tracheoesophageal groove, cardiac diaphragm angle, abdominal lymph node long diameter ≥ 5 mm or central lymph node necrosis. CT showed 3 or more lymph node aggregation or fusion or unclear boundary between lymph node and extranodal tissue. PET-CT showed that the SUV value of the lesion was > 2.5. According to the criteria, the lymph node region was divided to define the recurrence pattern: supraclavicular lymph node metastasis (area 1); mediastinal lymph node metastasis (2, 4 ~ 10, 15); upper abdominal lymph node (16 ~ 20 area) metastasis; anastomotic recurrence. These events were observed over a period of two years.

Clinical samples collection and cfDNA preparation

Patients with ESCC had eight milliliters of peripheral blood drawn into cfDNA Collection Tubes (Roche) and processed in one day. The plasma was separated by spinning at 13,500 × g for 12 min and then centrifuged at 1350 × g for 12 min, with both spins kept at 4 °C. Promptly, the plasma samples were frozen at −80 °C. cfDNA was successively isolated from the plasma by utilizing the Quick-cfDNA Serum & Plasma Kit (ZYMO) and thereafter preserved at a temperature of −80 °C. Nucleic acid electrophoresis was used to determine the fragment size of all cfDNA samples before the library was constructed.

5hmC library construction and high-throughput sequencing

Construction of 5hmC libraries utilized the efficient hmC-Seal technology [30]. The chemical labeling process, known for its sensitivity, allowed for cfDNA inputs as minimal as 1–10 ng. We used the KAPA Hyper Prep Kit (KAPA Biosystems) to 3′-adenylated cfDNA, followed next-generation sequencing protocols for end-repair, and ligated it using adapters compatible with Illumina. Glycosylation was carried out by subjecting ligated cfDNA to a 25 μL mixture at 37 °C for 2 h. This mixture included 50 mM HEPES buffer (pH 8.0), 25 mM MgCl2, 100 μM UDP6-N3-Glc, and 1 μM β-glucosyltransferase (NEB). Next, the DNA Clean & Concentrator Kit (ZYMO) was used to carry out the purification stages. The DNA that had been purified was subjected to an additional round of purification after being treated with 1 μL of DBCO-PEG4-biotin (4.5 mM in DMSO; Click Chemistry Tools) at 37 °C for 2 h. Simultaneously, 2.5 μL of streptavidin beads (Life Technologies) in 1 × buffer (5 mM Tris pH 7.5, 0.5 mM EDTA, 1 M NaCl, and 0.2% Tween 20) were introduced and left to incubate for 30 min. The beads were then resuspended in RNase-free water to prepare them for PCR, which involved fourteen to sixteen cycles after eight washes of the buffer. Following the instructions provided by Beckman, AMPure XP beads were used to purify the PCR products. Next, we used a Qubit 3.0 fluorometer from Life Technologies to assess the library concentration. Then, we sequenced the samples using the NextSeq 500 platform, and each read was 39 base pairs long.

Mapping and identifying 5hmC-enriched regions

Quality of the sequence was evaluated with the help of FastQC (version 0.11.5). The raw readings were aligned to the human genome (hg19) using bowtie2 (2.2.9) [31]. To guarantee that only unique, non-duplicate matches were found, SAMtools (version 1.3.1) [32] was used, with options SAMtools view-f2-F1548-q30 and SAMtools rmdup. The Integrated Genomics Viewer was used to display the converted BedGraph formatted, normalized extended pair-end reads using bedtools (version 2.19.1) [33] and finally turned into bigwig format by bedGraphToBigWig. The macs14-p1e-3-fBAMghs parameters were used in MACS (version 1.4.2) to identify possible 5hmC enriched regions (hMRs) [34]. Peak regions appearing in over ten samples and under 1000 bp were consolidated using bedtools merge. Blacklisted regions known for artifact signals, as identified by ENCODE, were excluded. hMRs for each patient were compiled by cross-referencing individual peak files with a unified peak file, excluding hMRs from chromosomes X and Y to refine downstream analysis. Subread version v1.5.3's Feature Counts was utilized to tally overlaps with genomic features [35]. Genomic annotations for hMRs were conducted using HOMER (version v4.10) [36], and all paired-end reads were normalized and transformed into bedgraph format with bam2bedgraph (version 1.0.4) [37].The genome-wide distribution of 5hmC was displayed using the Integrated Genomics Viewer (IGV) version 2.5.3 [38, 39]. The metagene profile was developed with ngsplot (version 2.61). The calculation of 5hmC

FPKM for hMRs involved using fragment counts from each hMR region obtained through bedtools [40].

Enrichment analysis

Using the DESeq2 package (v1.30.0) in R (version 4.3.0) [41, 42], areas with differential 5hmC enrichment were identified after excluding genes situated on chromosomes X and Y. With a log2foldchange larger than 0.25 and a P-value less than 0.05, these differentially 5hmC enriched regions (DhMRs) were identified for every pairwise comparison of groups. Comparisons were specifically conducted between the non-metastasis and metastasis groups, as well as between the non-recurrence and recurrence groups. The Pheatmap package (version 1.8.0) in R was used to run unsupervised hierarchical clustering and heatmap analyses. The DAVID web site, which uses hypergeometric tests for gene and protein functional annotation, also ran functional and pathway enrichment studies on genes with changed 5hmC alterations. The principal five KEGG pathways were subsequently highlighted to delineate the up- and down-regulated genes.

Feature selection, model training, and validation

Both the training and validation groups of patients with ESCC were randomly assigned. We used the train_test_split function in Python 3.6.10, which is part of Scikit-Learn (version 0.22.1) [43], to build a prediction model based on the logistic regression CV (LR) model. The DESeq2 package was used to identify DhMRs in the training cohort, with $p < 0.01$ and |log2FoldChange| greater than or equal to 0.25. As a precaution against overfitting, we used Scikit Learn's Statistical RegressionCV with the following parameters: estimator = LogisticRegressionCV (class_weight = 'balanced', cv = 2, max_iter = 1000), scoring = 'accuracy') to conduct five rounds of tenfold cross-validation. To create the final prediction model, cross-validation was painstakingly executed 100 times each round. Only markers that were present in three or more rounds were used. In order to predict the likelihood of metastasis and recurrence in the validation cohort, the LR model was trained using a set of chosen DhMR features (parameters: maxiter = 100, method = "lbfgs"). Receiver Operating Characteristics (ROC) analysis was used to evaluate the effectiveness of the model.

Development of weighted prediction score for ESCC metastasis and recurrence

A weighted prediction score (wp-score) was computed as the sum of the products of logistic model coefficients and corresponding 5hmC marker values for each individual, $wp - score = \sum_{k=1}^{n} \beta_k \times gene_k$, where β_k is the coefficient from the logistic model for the kth marker gene, and $gene_k$ represents the 5hmC level of that gene.

Statistical analysis

Continuous variables were presented as mean ± SD. Non-continuous and categorical variables were shown as frequencies or percentages and compared using the χ2 test. A two-sided P value of < 0.05 was considered statistically significant. These analyses were conducted using SPSS version 23.0. Survival outcomes of the groups were delineated through Kaplan–Meier survival analysis and the log-rank test, which provided survival curves and survival status descriptions.

Results

Clinical characteristics of ESCC patients

This study encompassed 122 individuals diagnosed with ESCC. Table 1 encapsulates the clinical profiles of these subjects, comprising 102 males and 20 females, whose ages spanned from 39 to 79 years. The median age was determined to be 62 years, resulting in a division of the cohort into 49 individuals younger than 60 and 73 individuals aged 60 or older. A total of 72 patients showed lymph node involvement, whereas 50 did not present any signs of metastasis. The distribution of TNM classifications revealed 37 instances at stage I, 18 at stage II, 55 at stage III, and 12 at stage IV. Primary tumors were located in the cervical segment (9 cases), upper thoracic (12 cases), middle thoracic (48 cases), and lower thoracic segments (53 cases). Differentiation of tumors was rated as high in 6 cases, moderate in 68 cases, low in 31 cases, and indeterminate in 17 cases. There were 72 patients who received chemotherapy and chemoradiation, including 61 cases of chemotherapy and 11 cases of chemoradiation. The chemotherapy regimens are listed in Table 1. Another 50 patients did not receive chemotherapy or chemoradiation. 105 patients received surgical treatment; 17 patients did not receive surgical treatment; during the follow-up period of two years, 30 patients had recurrence and 92 patients had no recurrence. A total of 22 patients had died by the follow-up date. For every patient, the expression of eight traditional tumor markers in the serum was detected. (Table S1).

Landscape of 5hmC profiles in cfDNA from patients with metastasis and recurrence of ESCC, differential gene analysis and function exploration

Subsequent analyses pinpointed regions enriched with 5hmC in each specimen, highlighting notable disparities in the 5hmC concentration of peripheral blood cfDNA across gene body among patients with and without metastasis and recurrence of ESCC (Fig. 1A). The average level of 5hmC in the recurrent group was the highest, followed by the metastatic group, the non-recurrent group, and the non-metastatic group. The variation trend

Table 1 Clinical characteristics of ESCC patients

Characteristics	Variables	Total (n (%))	Non-Metastasis (n (%))	Metastasis (n (%))	χ2	P	Non-Recur (n (%))	Recur (n (%))	χ2	P
Gender	Male	102(83.6%)	41(40.2%)	61(59.8%)	0.160	0.690	76(74.5%)	26(25.5%)	0.272	0.602
	Female	20(16.4%)	9(45.0%)	11(55.0%)			16(80.0%)	4(20.0%)		
Age	≤ 60	53(43.4%)	14(26.4%)	39(73.6%)	8.223	0.004	36(67.9%)	17(32.1%)	2.831	0.092
	> 60	69(56.6%)	36(52.2%)	33(47.8%)			56(81.2%)	13(18.8%)		
Smoker	No	35(28.7%)	19(54.3%)	16(45.7%)	3.591	0.058	30(85.7%)	5(14.3%)	2.810	0.094
	Yes	87(71.3%)	31(35.6%)	56(64.4%)			62(71.3%)	25(28.7%)		
Drinker	No	40(32.8%)	19(47.5%)	21(52.5%)	1.045	0.307	30(75.0%)	10(25.0%)	0.005	0.941
	Yes	82(67.2%)	31(37.8%)	51(62.2%)			62(75.6%)	20(24.4%)		
T stage	T0	9(7.4%)	6(66.7%)	3(33.3%)	24.222	0.000	8(88.9%)	1(11.1%)	13.987	0.016
	Tis	2(1.6%)	1(50.0%)	1(50.0%)			1(50.0%)	1(50.0%)		
	T1	34(27.9%)	24(70.6%)	10(29.4%)			29(85.3%)	5(14.7%)		
	T2	17(13.9%)	5(29.4%)	12(70.6%)			8(47.1%)	9(52.9%)		
	T3	57(46.7%)	14(24.6%)	43(75.4%)			45(78.9%)	12(21.1%)		
	T4	3(2.5%)	0(0.0%)	3(100.0%)			1(33.3%)	2(66.7%)		
N stage	N0	50(41.0%)	50(100.0%)	0(0.0%)	122.000	0.000	43(86.0%)	7(14.0%)	6.068	0.108
	N1	45(36.9%)	0(0.0%)	45(100.0%)			31(68.9%)	14(31.1%)		
	N2	18(14.8%)	0(0.0%)	18(100.0%)			13(72.2%)	5(27.8%)		
	N3	9(7.4%)	0(0.0%)	9(100.0%)			5(55.6%)	4(44.4%)		
Stage	Stage I	37(30.3%)	36(97.3%)	1(2.7%)	105.115	0.000	30(81.1%)	7(18.9%)	8.938	0.030
	Stage II	18(14.8%)	14(77.8%)	4(22.2%)			17(94.4%)	1(5.6%)		
	Stage III	55(45.1%)	0(0.0%)	55(100.0%)			39(70.9%)	16(29.1%)		
	Stage IV	12(9.8%)	0(0.0%)	12(100.0%)			6(50.0%)	6(50.0%)		
Location	Neck	9(7.4%)	0(0.0%)	9(100.0%)	10.220	0.017	3(33.3%)	6(66.7%)	10.389	0.016
	Upper	12(9.8%)	4(33.3%)	8(66.7%)			8(66.7%)	4(33.3%)		
	Middle	48(39.3%)	26(54.2%)	22(45.8%)			38(79.2%)	10(20.8%)		
	Lower	53(43.4%)	20(37.7%)	33(62.3%)			43(81.1%)	10(18.9%)		
Grade	GX	17(13.9%)	8(47.1%)	9(52.9%)	1.515	0.679	14(82.4%)	3(17.6%)	0.720	0.868
	G1	6(4.9%)	3(50.0%)	3(50.0%)			4(66.7%)	2(33.3%)		
	G2	68(55.7%)	29(42.6%)	39(57.4%)			51(75.0%)	17(25.0%)		
	G3	31(25.4%)	10(32.3%)	21(67.7%)			23(74.2%)	8(25.8%)		
Death	No	100(82.0%)	46(46.0%)	54(54.0%)	5.769	0.016	82(82.0%)	18(18.0%)	12.988	0.000
	Yes	22(18.0%)	4(18.2%)	18(81.8%)			10(45.5%)	12(54.5%)		
Regimen	No	50(41.0%)	22(44.0%)	28(56.0%)	4.53	0.806	44(88.0%)	6(12.0%)	16.989	0.03
	TP	54(44.3%)	20(37.0%)	34(63.0%)			34(63.0%)	20(37.0%)		
	CRT	10(8.2%)	5(50.0%)	5(50.0%)			8(80.0%)	2(20.0%)		
	TP + RT	1(0.8%)	0(0.0%)	1(100.0%)			0(0.0%)	1(100.0%)		
	TP + FOLFIRI	1(0.8%)	0(0.0%)	1(100.0%)			0(0.0%)	1(100.0%)		
	TP + CAPOX	1(0.8%)	0(0.0%)	1(100.0%)			1(100.0%)	0(0.0%)		
	DP	2(1.6%)	1(50.0%)	1(50.0%)			2(100.0%)	0(0.0%)		
	GP	2(1.6%)	1(50.0%)	1(50.0%)			2(100.0%)	0(0.0%)		
	PTX	1(0.8%)	1(100.0%)	0(0.0%)			1(100.0%)	0(0.0%)		
Surgery	No	17(13.9%)	0(0.0%)	17(100.0%)	13.717	0.000	12(70.6%)	5(29.4%)	0.248	0.619
	Yes	105(86.1%)	50(47.6%)	55(52.4%)			80(76.2%)	25(23.8%)		

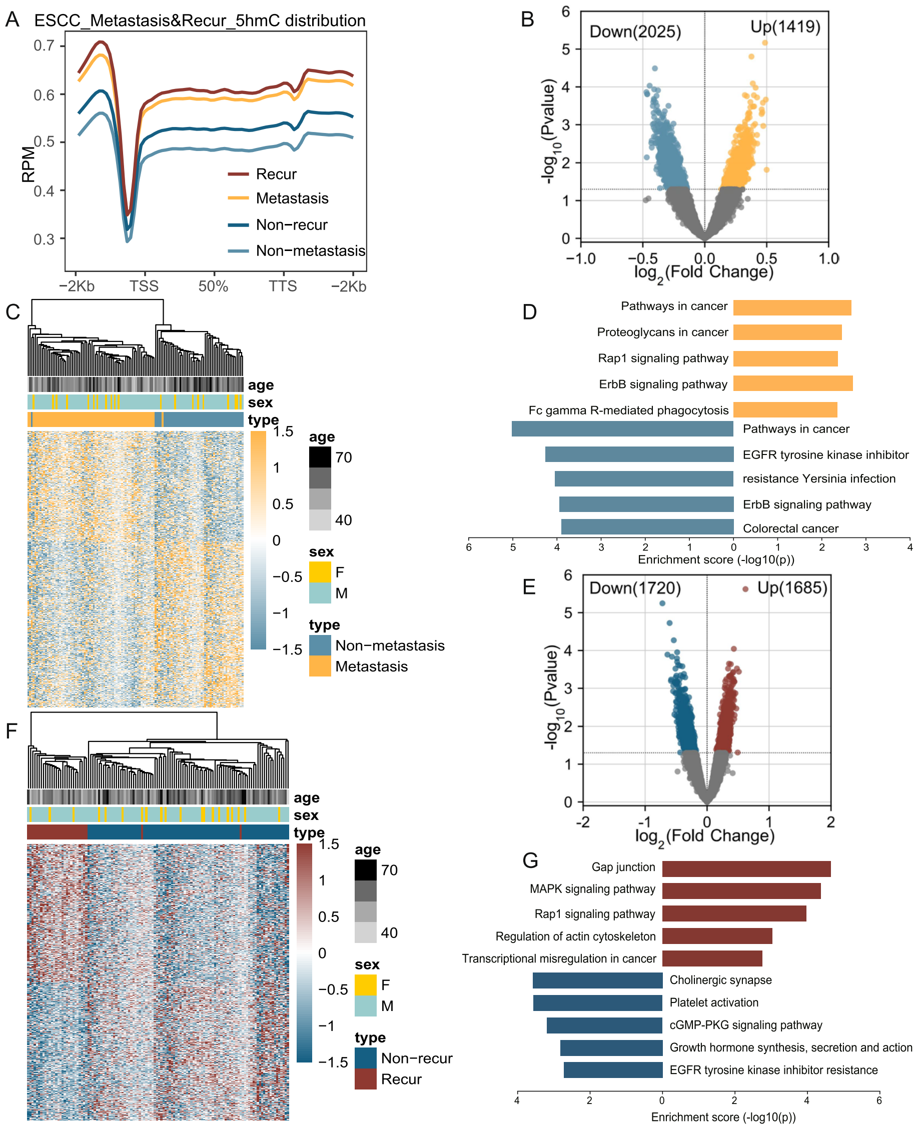


Fig. 1 illustrates the 5hmC profile landscape in cfDNA from ESCC patients with metastasis and recurrence, including the analysis of differential genes and exploration of their functions. **A** Shows the differentiated 5hmC read distribution among five groups. **B** Displays a volcano plot of differentially hydroxymethylated modification genes distribution in metastatic patients, genes with increased levels of hydroxymethylation modification marked in yellow and decreased levels of hydroxymethylation modification in blue. **C** A heatmap clusters the metastasis and non-metastasis samples using 600 differentially modified gene bodies identified from the overall samples. **D** KEGG analysis of differentially hydroxymethylated modification genes in metastatic patients indicated yellow for genes with increased levels of hydroxymethylation modification and blue for genes with decreased levels of hydroxymethylation modification. **E** Highlights a volcano plot of differentially hydroxymethylated modification genes distribution in recurrent patients, where red denotes genes with increased levels of hydroxymethylation modification and blue denotes genes with decreased levels of hydroxymethylation modification. **F** A heatmap clusters recurrence and non-recurrence samples using 600 differential hydroxymethylated modified genes identified from all samples. **G** KEGG analysis for differentially hydroxymethylated modification genes in recurrent patients showed red for genes with increased levels of hydroxymethylation modification and blue for genes with decreased levels of hydroxymethylation modification

of 5hmC in recurred patients and metastatic patients was similar in the gene body and its vicinity; that is, the content of 5hmC near the transcription start sites (TSS) decreased significantly, showing a clear valley shape. Genome-wide analysis of 5hmC-enriched regions in the metastatic and non-metastatic groups, recurrent and non-recurrent groups revealed that 5hmC-enriched regions were mostly enriched in introns, intergenic and promoter regions (Figure S3), and has a small valley at 2 kb downstream of the transcription termination site

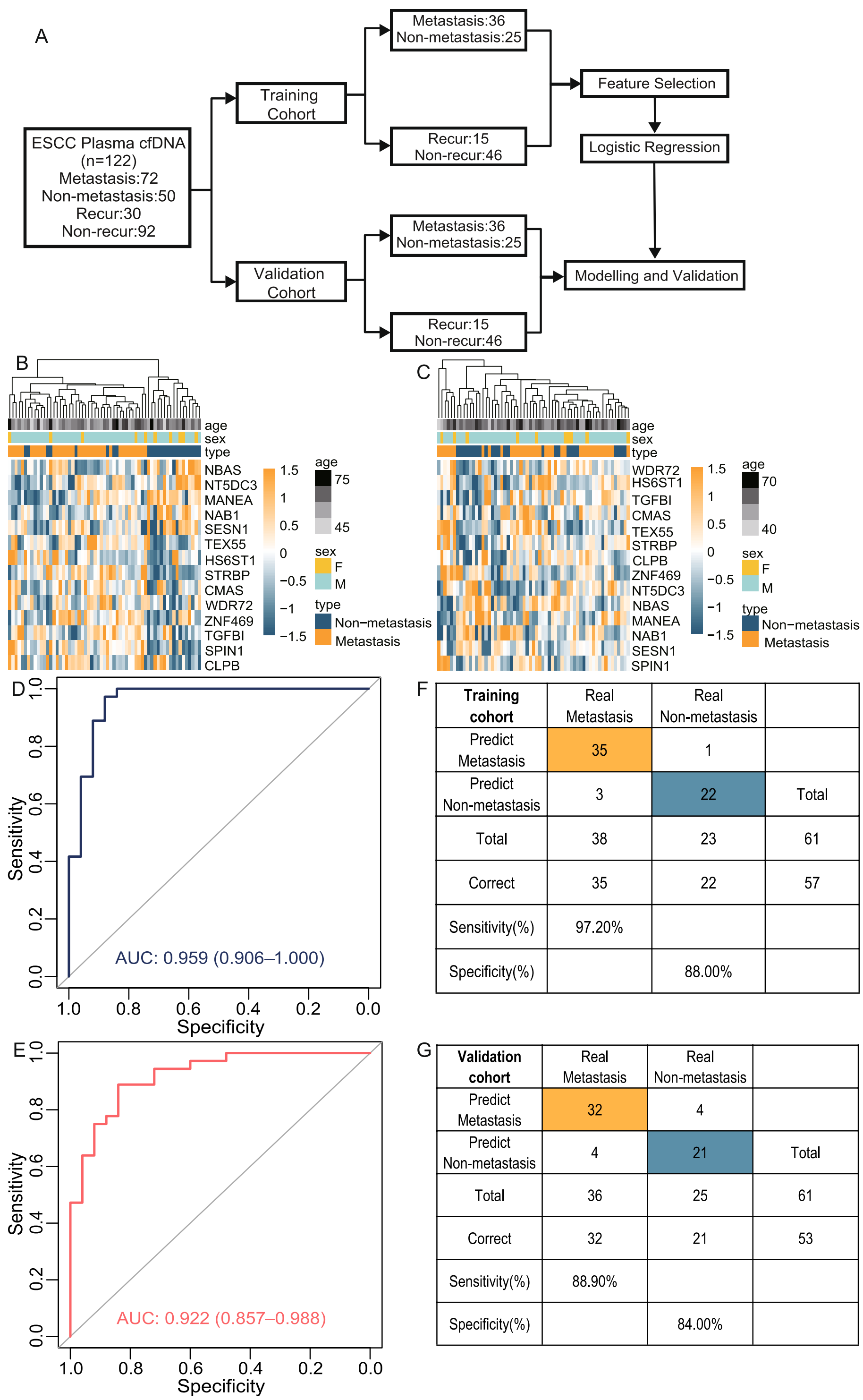
(TTS), which is similar to the results of previous studies. Then, the difference between metastatic and non-metastatic patients was analyzed. It was found that there were 1419 sites with 5hmC levels increased and 2025 sites with 5hmC levels decreased in metastatic patients (Fig. 1B and Table S2). Next, unsupervised clustering analysis was performed on the top 600 differential hydroxymethylated modified genes (DhMGs) between metastatic and non-metastatic patients to generate a heat map. Blue showed genes with decreased levels of hydroxymethylation modification, and yellow showed genes with increased levels of hydroxymethylation modification. The left side of the clustering tree is the metastatic sample, and the right side of the clustering tree is the non-metastatic sample (Fig. 1C). The classification trend of metastasis and non-metastasis is obvious. 5hmC differential modification sites can basically distinguish metastatic patients from non-metastatic patients. The analysis of differentially hydroxymethylated genes via pathway enrichment demonstrated that genes with increased levels of hydroxymethylation modification in metastatic cases primarily engaged in pathways associated with cancer, including proteoglycans in cancer, Rap1 signaling, ErbB signaling, and Fc gamma R-mediated phagocytosis. In contrast, the genes with decreased levels of hydroxymethylation modification showed a notable association with pathways related to cancer, resistance to EGFR tyrosine kinase inhibitors, Yersinia infection, ErbB signaling, and colorectal cancer (CRC) (Fig. 1D and Table S3). The differential analysis of recurrent and non-recurrent patients showed that there were 1685 sites with increased levels of 5hmC and 1720 sites with decreased levels of 5hmC in recurrent patients (Fig. 1E and Table S4). Then, unsupervised clustering analysis was performed on the top 600 DhMGs between recurrent and non-recurrent patients to generate heatmaps. Blue showed genes with decreased levels of hydroxymethylation modification, red showed genes with increased levels of hydroxymethylation modification, and the left side of the clustering tree was a recurrent sample, and the right side of the clustering tree was a non-recurrent sample (Fig. 1F). The trend of recurrence and non-recurrence classifications is obvious. The modifications in 5hmC served to distinguish effectively between recurrent and non-recurrent cases. The genes exhibiting increased level of hydroxymethylation modification in recurrent patients were primarily associated with gap junctions, MAPK pathways, Rap1 signaling, the regulation of the actin cytoskeleton, and transcriptional dysregulation related to cancer. Conversely, the genes that were decreased levels of hydroxymethylation modification primarily participated in processes related to the cholinergic synapse, platelet activation, the cGMP-PKG pathway, and the regulatory pathways governing growth hormone synthesis, secretion, and action, as well as resistance to EGFR tyrosine kinase inhibitors (Fig. 1G and Table S5).

Predictive models for detection metastasis of ESCC by 5hmC markers in cfDNA

Subjects with ESCC were randomly divided into training (36 exhibiting metastasis and 25 without) and validation cohorts (36 exhibiting metastasis and 25 without). A predictive logistic regression model based on 5hmC was developed within the training group to estimate the likelihood of metastasis in the validation set (Fig. 2A). Differential analysis ($|\log2FoldChange| \geq 0.25$, $p < 0.01$) identified 597 differential hydroxymethylation modified genes (DhMGs), comprising 242 hydroxymethylation modification levels increased and 355 hydroxymethylation modification levels decreased among those with metastasis compared to their non-metastatic counterparts. The application of a recursive feature elimination algorithm in conjunction with the logistic regression cross-validation estimator facilitated the reduction of 5hmC markers from 597 to 14, thereby optimizing the cross-validation performance. The logistic regression model identified 14 5hmC markers that effectively distinguished between metastatic and non-metastatic patients both in training (Fig. 2B) and validation cohorts (Fig. 2C). This study revealed that 14 5hmC markers were effective in forecasting both metastatic and non-metastatic outcomes within the training (AUC = 0.959) (Fig. 2D) and validation cohorts (AUC = 0.922) (Fig. 2E), achieving a sensitivity of 97.2% and a specificity of 88% in the training group (Fig. 2F), and a sensitivity of 88.9% and a specificity of 84.0% in the validation group (Fig. 2G). The findings suggest that variations in hydroxylmethylation levels of specific genes could act as indicators for the emergence of metastasis in individuals diagnosed with ESCC.

The correlation between the metastasis prediction score and clinical features

A primary aim was the creation of a robust, cohesive predictive model using 5hmC profiles in cfDNA to evaluate the risk of metastasis in ESCC patients. Consequently, a weighted prediction score, termed wp-scores, was generated for patients undergoing metastasis, which was significantly elevated in these patients as compared to non-metastatic individuals (Fig. 3A). The scores increased in correlation with the depth of tumor invasion, the quantity of lymph node metastases (LNM), and the clinical stage (Fig. 3B-D). Utilizing maxstat, the optimal cutoff value for the risk score was determined, establishing a minimum threshold of 25% and a maximum threshold of 75% of samples per group. Following this, patients were categorized into high- and low-risk groups



Training cohort	Real Metastasis	Real Non-metastasis	
Predict Metastasis	35	1	
Predict Non-metastasis	3	22	Total
Total	38	23	61
Correct	35	22	57
Sensitivity(%)	97.20%		
Specificity(%)		88.00%	

Validation cohort	Real Metastasis	Real Non-metastasis	
Predict Metastasis	32	4	
Predict Non-metastasis	4	21	Total
Total	36	25	61
Correct	32	21	53
Sensitivity(%)	88.90%		
Specificity(%)		84.00%	

Fig. 2 Predictive models for detection metastasis of ESCC by 5hmC biomarkers in cfDNA. **A** Schematic of the machine learning process. **B** Heatmaps illustrating the distribution of the 14 5hmC markers in training cohorts across patients exhibiting either metastasis or non-metastasis, with demographic data such as age and sex included. **C** Heatmaps illustrating the distribution of the 14 5hmC markers in validation cohorts across patients exhibiting either metastasis or non-metastasis, with demographic data such as age and sex included. **D** ROC curves of the metastasis classification system applying the 14 5hmC markers in training cohorts, depicting the true positive rate (sensitivity) against the false positive rate (1-specificity). **E** ROC curves of the metastasis classification system applying the 14 5hmC markers in validation cohorts, depicting the true positive rate (sensitivity) against the false positive rate (1-specificity). (**F**-**G**) Confusion matrices from the metastasis prediction model for each cohort

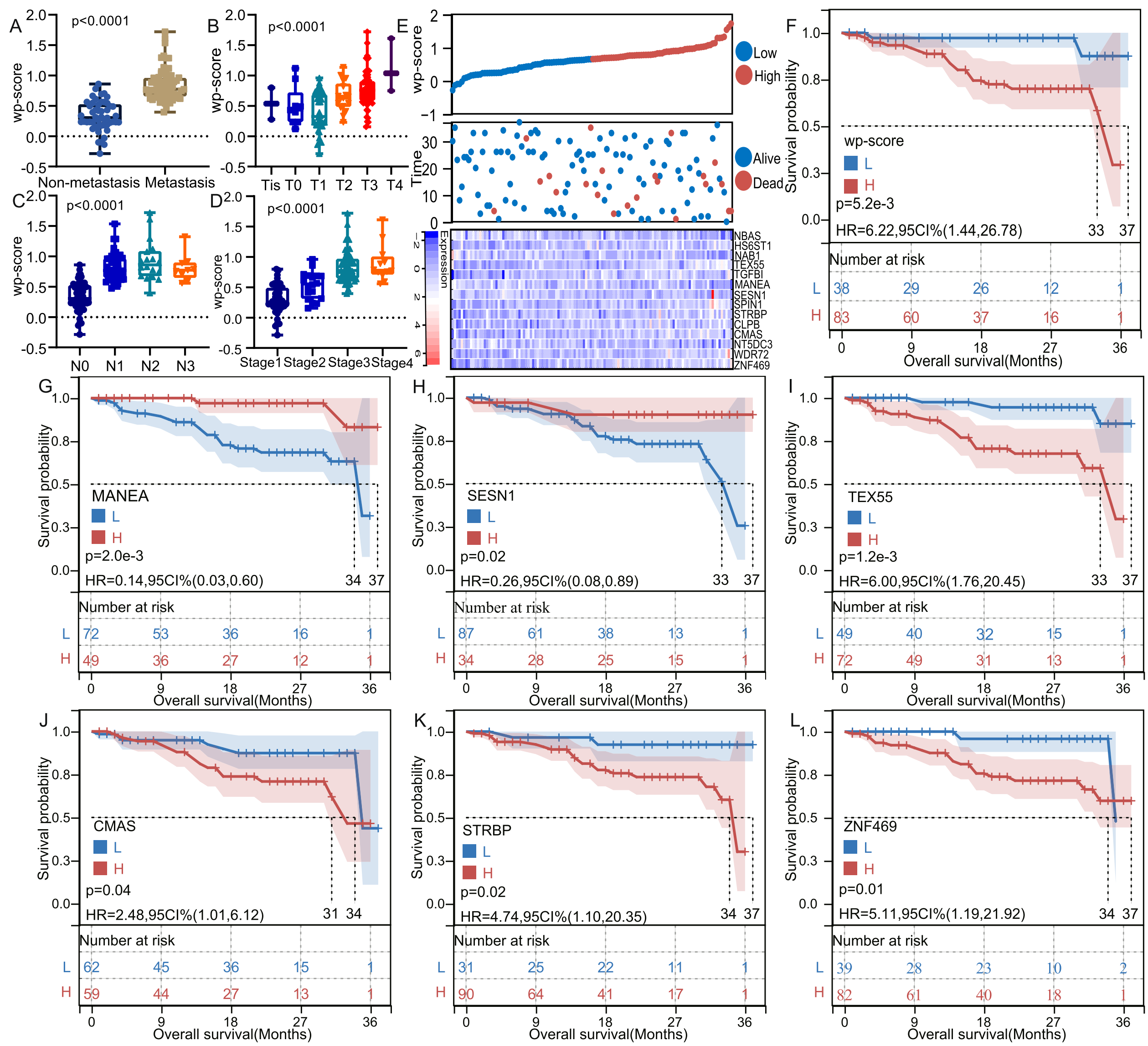


Fig. 3 The correlation between the metastasis prediction score and clinical features. **A** Differences in wp-scores based on 14 5hmC biomarkers between metastatic and non-metastatic ESCC patients. **B**-**D** Discrepancies in wp-scores associated with the depth of tumor invasion, the count of LNM, and clinical stage. **E** Profiles of risk scores, survival statuses, and hydroxymethylation modification level for the 14-gene set. **F** Prognostic differences according to the wp-score among ESCC patients. **G**-**L** Survival curves for ESCC patients categorized by increased or decreased levels of hydroxymethylation modifications in *TEX55*, *MANEA*, *CMAS*, *STRBP*, *SESN1*, and *ZNF469*, displaying overall survival (OS) time (months) versus survival probability

according to this threshold. The prognostic differences among these groups were evaluated utilizing the survfit function from the package survival, with the significance of these differences assessed through the logrank test, indicating a notable prognostic difference ($p=5.2\text{e-}3$) (Fig. 3F). The observed distribution of risk scores, survival statuses, and hydroxymethylation modification level associated with the 14-gene classifier (Fig. 3E) indicated that individuals exhibiting elevated risk scores experienced less favorable survival outcomes. Among these 14 genes, we observed significant prognostic differences between patients with hydroxymethylation modification level increased of *MANEA* (Fig. 3G) and *SESN1* (Fig. 3H) and those with hydroxymethylation modification level decreased of *TEX55* (F ig. 3I), *CMAS* (Fig. 3J), *STRBP* (Fig. 3K), and *ZNF469* (Fig. 3L).

Predictive models for detection recurrence of ESCC by 5hmC biomarkers in cfDNA

Individuals diagnosed with ESCC were categorized into a training cohort consisting of 15 individuals with recurrence and 46 without, alongside a validation cohort comprising 15 with recurrence and 46 without. In the training group, a logistic regression model based on 5hmC was developed to predict recurrence (Fig. 2A). The initial differential analysis ($|\log2FoldChange| \geq 0.25$, $p < 0.01$) identified 592 differential hydroxymethylation modified genes (DhMGs), comprising 293 hydroxymethylation modification levels increased and 299 hydroxymethylation modification levels decreased in patients with recurrence compared to those without. The recursive feature elimination method associated with the logistic regression cross-validation estimator was subsequently employed, reducing the quantity of 5hmC markers from 592 to 11, thus optimizing the cross-validation score. The selected 11 5hmC markers demonstrated the ability to distinguish between recurrence and non-recurrence across both training (Fig. 4A) and validation cohorts (Fig. 4B). They effectively predicted recurrence and non-recurrence in the training cohort (AUC = 0.981) (Fig. 4C) and the validation cohort (AUC = 0.936) (Fig. 4D), achieving 100% sensitivity and 91.3% specificity in the training group (Fig. 4E), and 93.3% sensitivity and 89.1% specificity in the validation group (Fig. 4F). The results suggest that alterations in hydroxylmethylation levels of particular genes could serve as indicators for the likelihood of recurrence in patients with ESCC.

The correlation between the recurrence prediction score and clinical features

The primary objective was to develop a cohesive and effective predictive model utilizing 5hmC profiles in cfDNA to evaluate the likelihood of recurrence in patients with ESCC. In pursuit of this objective, we established a weighted prediction score, revealing that wp-scores derived from 11 5hmC biomarkers were markedly elevated in patients demonstrating recurrence compared to those who did not (Fig. 5A). The maxstat was utilized to determine the optimal cutoff value for the risk score, ensuring that the minimum group sample size exceeded 25% while the maximum remained under 75%. Subsequently, patients were categorized into high- and low-risk groups based on this threshold. We conducted a detailed examination of the prognostic distinctions between these groups utilizing the survfit function from the package survival. The significance of these distinctions was evaluated through the logrank test, which indicated a significant prognostic variance ($p = 3.5e{-3}$) (Fig. 5C). The evaluation of the risk score distribution, survival status, and hydroxymethylation modification level of the 11-gene classifier (Fig. 5B) indicated that individuals classified in the higher-risk group exhibited poorer survival outcomes relative to those in the lower-risk group. Among these 11 genes, we observed significant prognostic differences between patients with hydroxymethylation modification level increased of *LLPH* (Fig. 5D) and those with hydroxymethylation modification level decreased of *ANGPT2* (Fig. 5E) and *LINC02694* (Fig. 5F). We utilized TCGA data from UALCAN (http://ualcan.path.uab.edu/analysis.html) to confirm and graphically illustrate the expression patterns of the key 5hmC markers in our model. It was established that the expression levels of these 5hmC markers in esophageal cancer tissues differed from those in normal tissues (Figure S7A-G, Figure S10A-B). The markers TGFBI, STRBP, CMAS, CLPB, WDR72, ZNF469, HS6ST1, SERPINH1, and ANGPT2 were consistent with the observed changes in hydroxymethylation modification levels in our data (Figure S4A-E, H-I, Figure S6B, F). The expression levels of the 5hmC markers TGFBI, STRBP, CMAS, CLPB, WDR72, ZNF469, and HS6ST1 in esophageal cancer tissues at various clinical stages and with lymph node metastasis were significantly different from those in normal tissues (Figure S8A-G, Figure S9A-G). This finding aligns with the changes in hydroxymethylation modification levels observed in our data (Figure S5A-B). Additionally, the high expression of the 5hmC marker ANGPT2 in the recurrence model demonstrates a significant prognostic difference in patients with esophageal cancer (Figure S10C), which corresponds with the notable prognostic difference associated with the increase in ANGPT2 hydroxymethylation modification levels in our data (Fig. 5E).

5hmC biomarkers have higher prediction performance than traditional markers in detecting metastasis and recurrence

In predicting the metastasis and recurrence of ESCC patients, the model based on 5hmC biomarkers was significantly better than the model based on traditional tumor markers. Figure 6A shows the performance of traditional tumor markers in distinguishing the metastasis of ESCC patients in the validation cohort. The results showed that the AUCs of *CEA*, *CYFRA21-1*, *SCC*, *CA125*, *CA199*, *NSE*, *CA72.4*, and *CA242* were 0.517, 0.529, 0.528, 0.508, 0.478, 0.540, and 0.503, respectively. The predictive performance was lower than the 14 5hmC biomarkers we screened (Fig. 6B and 6C). Similarly, the traditional tumor markers obtained the same results in distinguishing the recurrence of ESCC patients in the validation cohort (Fig. 6D), and 11 5hmC biomarkers had higher predictive efficacy (Fig. 6E and 6F). Additionally, the prediction performance of each traditional marker

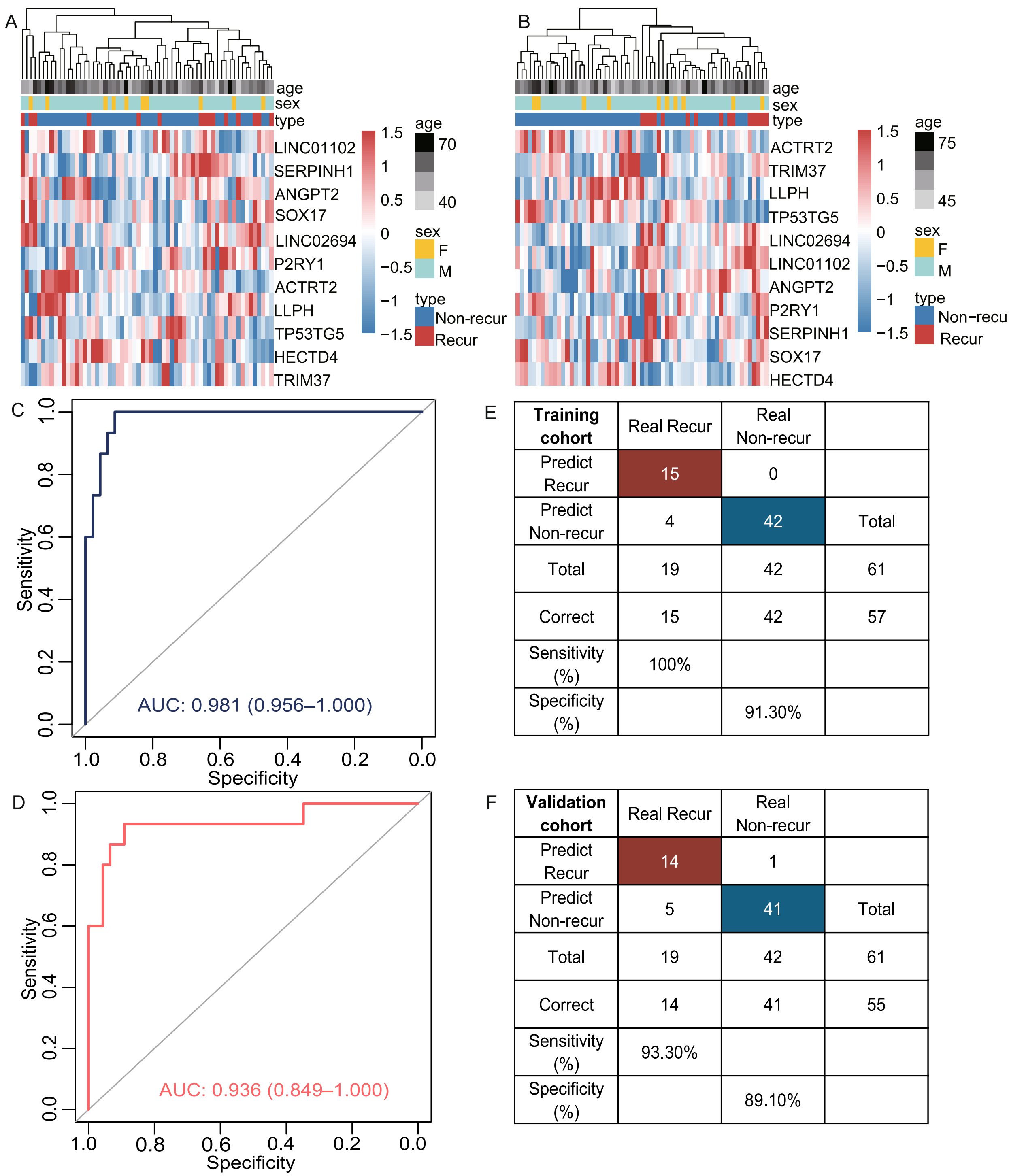


Fig. 4 Predictive models for detection recurrence of ESCC by 5hmC biomarkers in cfDNA. **A** Heatmaps showcasing the 11 5hmC markers for patients with and without recurrence in training cohorts, incorporating demographic details such as age and sex. **B** Heatmaps showcasing the 11 5hmC markers for patients with and without recurrence in validation cohorts, incorporating demographic details such as age and sex. **C** ROC curves of the recurrence classification model using 11 5hmC markers in training cohorts, plotting the true positive rate (sensitivity) against the false positive rate (1-specificity). **D** ROC curves of the recurrence classification model using 11 5hmC markers in validation cohorts, plotting the true positive rate (sensitivity) against the false positive rate (1-specificity). **E**–**F** Confusion matrices derived from the recurrence prediction model for each cohort

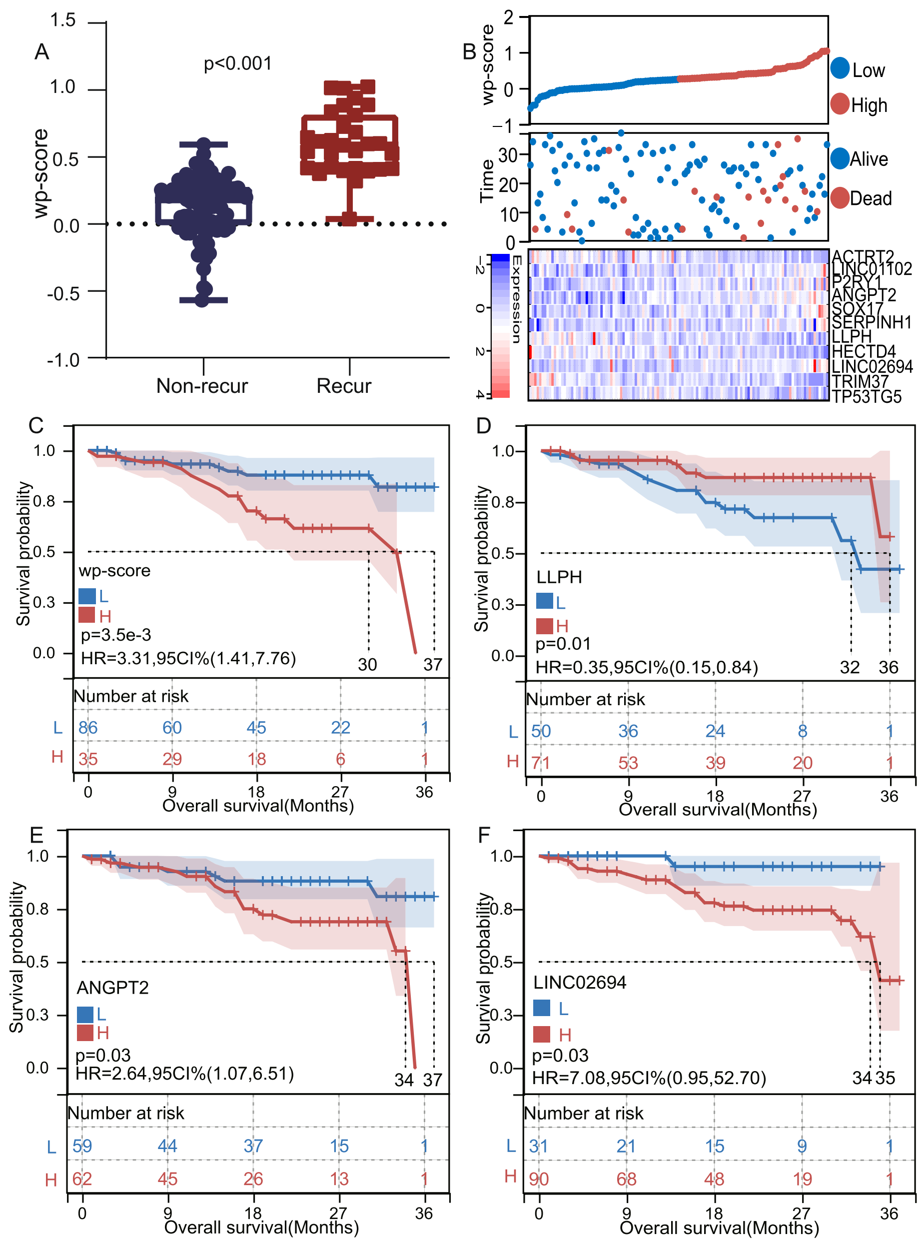


Fig. 5 The correlation between the recurrence prediction score and clinical features. **A** Differences in wp-scores based on 11 5hmC biomarkers between recurrence and non-recurrence ESCC patients. **B** Distributions of risk scores, survival statuses, and hydroxymethylation modification level of the 11 genes. **C** Variations in prognostic outcomes based on the wd-score among ESCC patients. **D**-**F** Survival curves for ESCC patients categorized by the increased or decreased levels of hydroxymethylation modifications in *LLPH*, *ANGPT2*, and *LINC02694*, with the x-axis indicating OS time (months) and the y-axis depicting the probability of survival

significantly improved when combined with 5hmC markers (Figure S1 and S2).

Discussion

ESCC is a deadly disease prone to recurrence and metastasis. At present, there are still unmet needs for biomarkers for the detection of metastasis and recurrence. 5hmC, a novel epigenetic marker, is essential in the modulation of gene expression and is associated with numerous biological processes, including cancer and metabolic disorders [44]. Despite the low levels of cfDNA 5hmC present in the bloodstream, their potential utility as biomarkers for a range of cancers is noteworthy [25]. Therefore, cfDNA 5hmC signal characteristics in blood could be

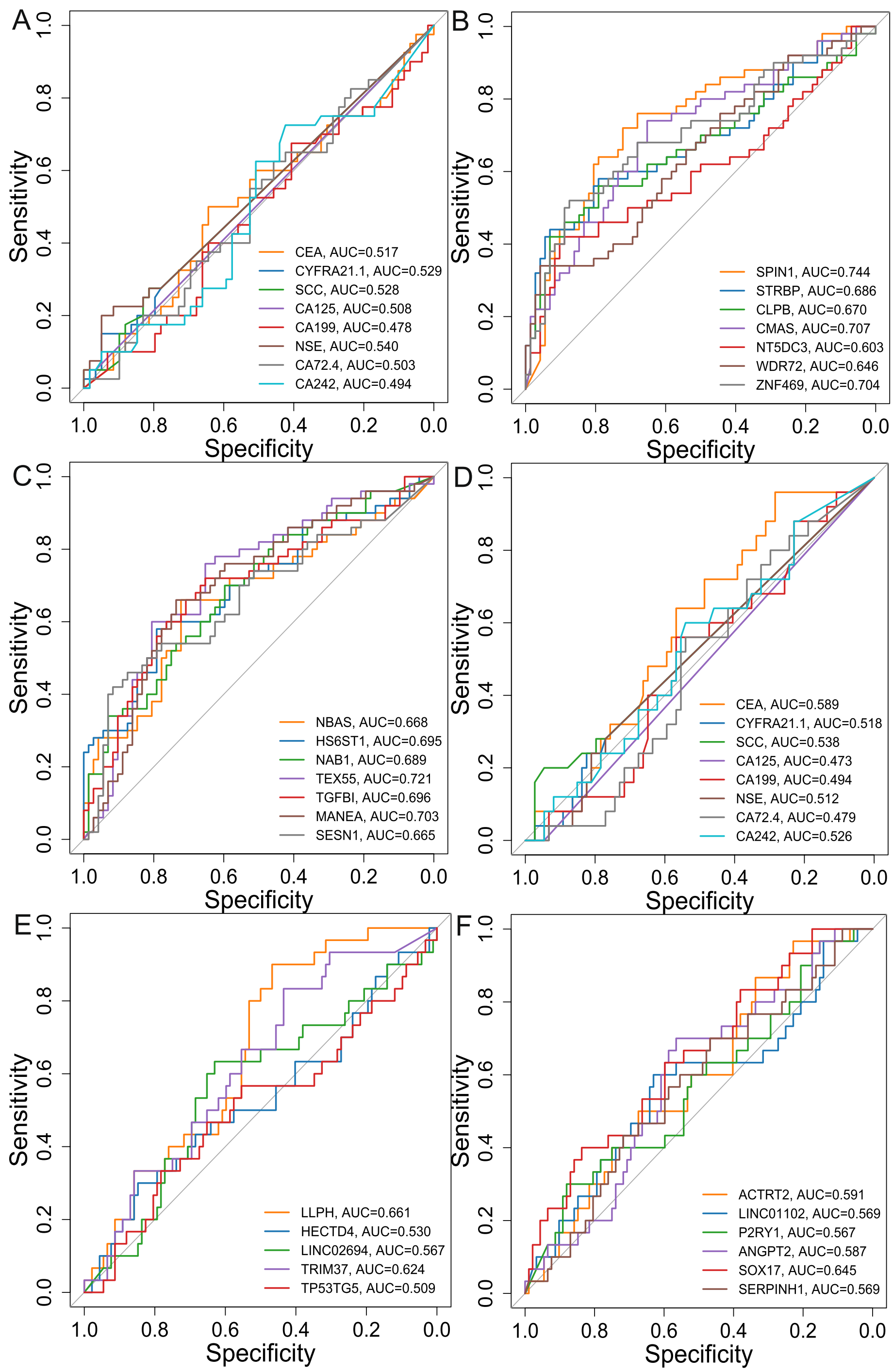


Fig. 6 5hmC biomarkers have higher prediction performance than traditional markers in detecting metastasis and recurrence. **A** ROC curve analyzing traditional tumor markers in the validation cohort of ESCC patients with metastasis. **B**-**C** ROC curves for 14 5hmC biomarkers in the validation cohort of ESCC patients with metastasis. **D** ROC curve examining traditional tumor markers in the validation cohort of ESCC patients with recurrence. **E**–**F** ROC curves for 14 5hmC biomarkers in the validation cohort of ESCC patients with recurrence, plotting the true positive rate (sensitivity) against the false positive rate (1-specificity)

robust indicators for a range of diseases. The 5hmC-Seal technique efficiently and specifically captures 5hmC, making it suitable for samples with minimal DNA and offering low sequencing costs that are ideal for large-scale testing. Based on liquid biopsy, this technique provides the advantages of being minimally invasive, non-invasive, and facilitating early screening, which enhances patient compliance and cost-effectiveness. It is particularly suitable for early diagnosis and screening. When combined with other omics data, it offers a comprehensive understanding of cancer biology, yielding more complete diagnostic and prognostic information. However, the costs associated with reagents and equipment are high, and multi-omics analysis further increases expenses. Additionally, the technique cannot perform single-base resolution analysis, and chemical capture methods have inherent limitations. The amount of circulating tumor DNA (ctDNA) in plasma samples is often limited, and the interpretation and standardization of data can be complex, necessitating professional support. Furthermore, consistency of data across different laboratories must be addressed. More large-scale randomized clinical trials are needed to verify the clinical effectiveness and reliability of this technique. Previous investigations have identified 5hmC modifications associated with esophageal cancer in plasma cfDNA, proposing that these biomarkers could aid in the early detection of the disease [45, 46]. In this study, we utilized the hmC-Seal sequencing technique to examine cfDNA 5hmC in individuals diagnosed with ESCC, aiming to uncover dependable biomarkers indicative of metastasis and recurrence.

Initially, we applied the sensitive hmC-Seal method [47] to generate comprehensive cfDNA 5hmC profiles for ESCC patients, both with and without metastasis and recurrence. Consistent and significantly elevated 5hmC signals were observed in the gene bodies and promoter regions of patients exhibiting metastasis and recurrence. Research has demonstrated that genes with elevated levels of 5hmC are closely linked to the initiation and advancement of cancer [48]. A significant disparity in 5hmC levels was identified between metastatic and non-metastatic patients, and differential hydroxymethylation modified genes were able to distinctly differentiate between these groups, suggesting that 5hmC markers could function as effective surveillance markers for metastasis. Similarly, we obtained the same results in recurred and non-recurred patients, indicating that 5hmC markers can also be used as a monitoring marker for recurrence.

Next, we performed functional analysis of these differential hydroxymethylation modified genes and found that these genes were mainly concentrated in pathways highly related to tumors. Pathway analysis of differentially modified 5hmC between metastatic and non-metastatic patients and between recurrent and non-recurrent patients suggested that both groups were also enriched in Rap1 signaling pathway. Ras-related protein 1 (Rap1), classified within the Ras superfamily, functions as a small G protein that plays a crucial role in cell signal transduction [49, 50]. The unusual activation of the Rap1 pathway plays a crucial role in tumor development by promoting the proliferation, migration, and invasion of cancer cells, thereby affecting both metastasis and recurrence [51–53]. Studies have repeatedly demonstrated that the disruption of this pathway is closely linked to tumor metastasis and recurrence, indicating its promise as a target for therapeutic intervention in managing cancer progression. The results indicate that 5hmC markers, especially those obtained from cfDNA, are intricately associated with the advancement of ESCC and function as reliable epigenetic indicators for monitoring both metastasis and recurrence.

Additionally, utilizing a machine learning-based tumor classifier, we pinpointed potential 5hmC-based markers within circulating cfDNA from ESCC patients experiencing metastasis and recurrence. The constructed predictive model, characterized by its high sensitivity and specificity, successfully differentiated between metastatic and non-metastatic as well as recurrent and non-recurrent patients, and delineated pronounced prognostic differences between high-risk and low-risk groups. Notably, 14 5hmC markers isolated through machine learning algorithms effectively distinguished between metastatic and non-metastatic patients in both the training and validation cohorts. Moreover, a logistic regression model utilizing these 14 markers achieved sensitivities of 0.889 and specificities of 0.840 (AUC = 0.922). In parallel, 11 5hmC markers similarly identified through machine learning distinctly separated recurrent from non-recurrent patients across both cohorts, with the corresponding logistic regression model demonstrating sensitivities of 0.933 and specificities of 0.891 (AUC = 0.936). The gathered results collectively indicate the potential of 5hmC markers derived from cfDNA as effective biomarkers for the non-invasive identification of ESCC metastasis and recurrence. These markers are associated with critical clinical parameters, such as the depth of tumor invasion, the number of LNM, and the clinical staging of the disease. Several genes among these markers exhibited significant prognostic disparities among groups, underscoring their potential role in monitoring metastasis and recurrence. Studies have revealed the involvement of the Toll-like receptor (TLR) pathway, PD-L1 expression, and the PD-1 checkpoint pathway as possible downstream targets of SESN1 in cancer, with SESN1 functioning as a novel tumor suppressor in neuroblastoma (NB) through the TLR pathway. High *SESN1* expression has been linked with elevated immune scores, highlighting its importance for the immunotherapy and prognosis of NB [54]. *ANGPT2*, recognized for its role in the regulation of development, progression, invasion, and metastasis in diverse malignant tumors, is proposed as a biomarker and therapeutic target for the early diagnosis of cancer. The elevated expression of ANGPT2 in non-small cell lung cancer (NSCLC) is associated with a negative prognosis, highlighting its significance as an important prognostic indicator [55]. Additionally, bioinformatics analyses have linked *ANGPT2* upregulation with improved OS in gastric cancer patients, suggesting its potential as a prognostic biomarker [56]. High *ANGPT2* expression levels in esophageal cancer are associated with adverse patient outcomes [57]. Moreover, from our results, 5hmC markers show higher predictive performance than traditional biomarkers, such as *CEA, CYFRA21-1, SCC, CA125, CA199, NSE, CA72.4,* and *CA242*.

Overall, our results demonstrate that 5hmC signals in cfDNA are effective as minimally invasive biomarkers for monitoring metastasis and recurrence in ESCC. Additionally, these markers can stratify patients with metastatic and recurrent ESCC into low-risk and high-risk categories, thereby guiding appropriate therapeutic strategies.

Limitations

There are several limitations to this study. First, lack of double-blind validation cohort. Due to the significant heterogeneity of tumors, forthcoming research should encompass multi-center and independent clinical trials to identify specific cell-free 5hmC markers for diseases. Future studies must focus on validating the sensitivity, specificity, and precision of these markers, discovering more dependable 5hmC markers, and exploring the relationship between 5hmC levels, various confounding factors, and their changes following treatment. These endeavors are essential to propel the clinical adoption of this innovative technology in the field of precision oncology. Second, our study only predicts metastasis, not the specific prediction of the metastasis site. Third, there is no longitudinal cohort. Fourth, the population we included was only Chinese patients, without considering the differences in ethnicity, which limits its applicability to diverse populations. Expanding future research to encompass multi-ethnic and geographically varied cohorts would strengthen the universal applicability of the findings.

Conclusions

In conclusion, our studies have shown that 5hmC markers from plasma cfDNA are capable of detecting both metastasis and recurrence in ESCC. These insights could lead to the development of novel methodologies for the surveillance of metastasis and recurrence in ESCC.

Abbreviations

ESCC	Esophageal squamous cell carcinoma
cfDNA	Cell-free DNA
5hmC	5-Hydroxymethylcytosine
DhMGs	Differential hydroxymethylation modified genes
hMRs	5HmC-enriched regions
RFECV	Recursive feature elimination algorithm
ROC	Receiver operating characteristic
AUC	The area under ROC curves
KEGG	Kyoto Encyclopedia of Genes and Genomes
TSS	Transcription start sites
TTS	Transcription termination site
OS	Overall survival

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12957-025-03747-9.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.
Supplementary Material 9.
Supplementary Material 10.
Supplementary Material 11.
Supplementary Material 12.
Supplementary Material 13.
Supplementary Material 14.
Supplementary Material 15.
Supplementary Material 16.

Acknowledgements

We thank all study participants, research staffs, and students who participated in this work.

Authors' contributions

HYC, LC, XZK and YCT conceived the research and designed the experiments. LZ, YZC, HTZ, ZL and GBX recruited patients, collected blood samples, and registered clinical information. XHL, BXZ, HX and SK accomplished the sequencing experiments and analyzed raw data. SK analyzed the data with the help of HYC. SK wrote and revised the manuscript based on suggestions from HYC, JL and YCT. JL and YCT offered funds for the research. All authors read and approved the final manuscript.

Funding

This work was supported by the Natural Science Foundation of China (82274034), the Xinjiang Key Laboratory of Natural Medicines Active Components and Drug Release Technology (XJDX1713), the Xinjiang Key Laboratory of Biopharmaceuticals and Medical Devices, and the Engineering

Research Center of Xinjiang and Central Asian Medicine Resources, Ministry of Education.

Data availability
Data is provided within the manuscript or supplementary information files

Declarations

Ethics approval and consent to participate
The study was conducted according to the guidelines of the Helsinki Declaration and was approved by the Ethics Committee of Peking University Cancer Hospital. Written informed consent was obtained from all participants.

Consent for publication
All authors contributed to the manuscript and approve its submission.

Competing interests
The authors declare no competing interests.

Received: 28 November 2024 Accepted: 7 March 2025
Published online: 15 March 2025

References
1. Arnold M, Soerjomataram I, Ferlay J, Forman D. Global incidence of oesophageal cancer by histological subtype in 2012. Gut. 2015;64(3):381–7.
2. Hamai Y, Hihara J, Emi M, Furukawa T, Ibuki Y, Yamakita I, et al. Treatment Outcomes and Prognostic Factors After Recurrence of Esophageal Squamous Cell carcinoma. World J Surg. 2018;42(7):2190–8.
3. Jiang L, Lin X, Chen F, Qin X, Yan Y, Ren L, et al. Current research status of tumor cell biomarker detection. Microsyst Nanoeng. 2023;9:123.
4. Das S, Dey MK, Devireddy R, Gartia MR. Biomarkers in Cancer Detection, Diagnosis, and Prognosis. Sensors (Basel). 2023;24(1):37.
5. Bai JW, Qiu SQ, Zhang GJ. Molecular and functional imaging in cancer-targeted therapy: current applications and future directions. Signal Transduct Target Ther. 2023;8(1):89.
6. Ali S, Fenerty S, Jonnalagadda P. Imaging Modalities in the Detection and Diagnosis of Metastatic Disease. In: Leong SP, Nathanson SD, Zager JS, editors. Cancer Metastasis Through the Lymphovascular System. Cham: Springer International Publishing; 2022. p. 283–93.
7. Sekiguchi M, Matsuda T. Limited usefulness of serum carcinoembryonic antigen and carbohydrate antigen 19–9 levels for gastrointestinal and whole-body cancer screening. Sci Rep. 2020;10(1):18202.
8. Yuan Z, Wang X, Geng X, Li Y, Mu J, Tan F, et al. Liquid biopsy for esophageal cancer: Is detection of circulating cell-free DNA as a biomarker feasible? Cancer Commun (Lond). 2021;41(1):3–15.
9. Labgaa I, Villanueva A, Dormond O, Demartines N, Melloul E. The Role of Liquid Biopsy in Hepatocellular Carcinoma Prognostication. Cancers (Basel). 2021;13(4):659.
10. Osumi H, Shinozaki E, Yamaguchi K, Zembutsu H. Clinical utility of circulating tumor DNA for colorectal cancer. Cancer Sci. 2019;110(4):1148–55.
11. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300–3.
12. Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet. 2011;13(1):7–13.
13. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930–5.
14. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer. 2003;3(4):253–66.
15. Chen HY, Zhang WL, Zhang L, Yang P, Li F, Yang ZR, et al. 5-Hydroxymethylcytosine profiles of cfDNA are highly predictive of R-CHOP treatment response in diffuse large B cell lymphoma patients. Clin Epigenetics. 2021;13(1):33.
16. Wang Z, Du M, Yuan Q, Guo Y, Hutchinson JN, Su L, et al. Epigenomic analysis of 5-hydroxymethylcytosine (5hmC) reveals novel DNA methylation markers for lung cancers. Neoplasia. 2020;22(3):154–61.
17. Tong M, Gao S, Qi W, Shi C, Qiu M, Yang F, et al. 5-Hydroxymethylcytosine as a potential epigenetic biomarker in papillary thyroid carcinoma. Oncol Lett. 2019;18(3):2304–9.
18. Thomson JP, Meehan RR. The application of genome-wide 5-hydroxymethylcytosine studies in cancer research. Epigenomics. 2017;9(1):77–91.
19. Zhang F, Liu Y, Zhang Z, Li J, Wan Y, Zhang L, et al. 5-hydroxymethylcytosine loss is associated with poor prognosis for patients with WHO grade II diffuse astrocytomas. Sci Rep. 2016;6:20882.
20. Chen K, Zhang J, Guo Z, Ma Q, Xu Z, Zhou Y, et al. Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer. Cell Res. 2016;26(1):103–18.
21. Mariani CJ, Madzo J, Moen EL, Yesilkanal A, Godley LA. Alterations of 5-hydroxymethylcytosine in human cancers. Cancers (Basel). 2013;5(3):786–814.
22. Orr BA, Haffner MC, Nelson WG, Yegnasubramanian S, Eberhart CG. Decreased 5-hydroxymethylcytosine is associated with neural progenitor phenotype in normal brain and shorter survival in malignant glioma. PLoS ONE. 2012;7(7): e41036.
23. Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, et al. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell. 2012;150(6):1135–46.
24. Cui XL, Nie J, Ku J, Dougherty U, West-Szymanski DC, Collin F, et al. A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. Nat Commun. 2020;11(1):6161.
25. Zeng C, Stroup EK, Zhang Z, Chiu BC, Zhang W. Towards precision medicine: advances in 5-hydroxymethylcytosine cancer biomarker discovery in liquid biopsy. Cancer Commun (Lond). 2019;39(1):12.
26. Gao P, Lin S, Cai M, Zhu Y, Song Y, Sui Y, et al. 5-Hydroxymethylcytosine profiling from genomic and cell-free DNA for colorectal cancers patients. J Cell Mol Med. 2019;23(5):3530–7.
27. Song CX, Yin S, Ma L, Wheeler A, Chen Y, Zhang Y, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. Cell Res. 2017;27(10):1231–42.
28. Li W, Zhang X, Lu X, You L, Song Y, Luo Z, et al. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. Cell Res. 2017;27(10):1243–57.
29. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform. 2014;48:193–204.
30. Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol. 2011;29(1):68–72.
31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
33. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinform. 2014;47:11.2.1-34.
34. Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447(7146):799–816.
35. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.
36. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576–89.
37. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.
38. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92.
39. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6.

40. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7): e47.
41. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
42. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.
43. Swami A, Jain R. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2013;12(10):2825–30.
44. Pinzón-Cortés JA, Perna-Chaux A, Rojas-Villamizar NS, Díaz-Basabe A, Polanía-Villanueva DC, Jácome MF, et al. Effect of diabetes status and hyperglycemia on global DNA methylation and hydroxymethylation. Endocr Connect. 2017;6(8):708–25.
45. Lu D, Wu X, Wu W, Wu S, Li H, Zhang Y, et al. Plasma cell-free DNA 5-hydroxymethylcytosine and whole-genome sequencing signatures for early detection of esophageal cancer. Cell Death Dis. 2023;14(12):843.
46. Tian X, Sun B, Chen C, Gao C, Zhang J, Lu X, et al. Circulating tumor DNA 5-hydroxymethylcytosine as a novel diagnostic biomarker for esophageal cancer. Cell Res. 2018;28(5):597–600.
47. Han D, Lu X, Shih AH, Nie J, You Q, Xu MM, et al. A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. Mol Cell. 2016;63(4):711–9.
48. Qi J, Shi Y, Tan Y, Zhang Q, Zhang J, Wang J, et al. Regional gain and global loss of 5-hydroxymethylcytosine coexist in genitourinary cancers and regulate different oncogenic pathways. Clin Epigenetics. 2022;14(1):117.
49. Looi CK, Hii LW, Ngai SC, Leong CO, Mai CW. The Role of Ras-Associated Protein 1 (Rap1) in Cancer: Bad Actor or Good Player? Biomedicines. 2020;8(9):337.
50. Huang M, Anand S, Murphy EA, Desgrosellier JS, Stupack DG, Shattil SJ, et al. EGFR-dependent pancreatic carcinoma cell metastasis through Rap1 activation. Oncogene. 2012;31(22):2783–93.
51. Zhou S, Liang Y, Zhang X, Liao L, Yang Y, Ouyang W, et al. SHARPIN Promotes Melanoma Progression via Rap1 Signaling Pathway. J Invest Dermatol. 2020;140(2):395-403.e6.
52. Li Q, Xu A, Chu Y, Chen T, Li H, Yao L, et al. Rap1A promotes esophageal squamous cell carcinoma metastasis through the AKT signaling pathway. Oncol Rep. 2019;42(5):1815–24.
53. Bailey CL, Kelly P, Casey PJ. Activation of Rap1 promotes prostate cancer metastasis. Cancer Res. 2009;69(12):4962–8.
54. Hua Z, Chen B, Gong B, Lin M, Ma Y, Li Z. SESN1 functions as a new tumor suppressor gene via Toll-like receptor signaling pathway in neuroblastoma. CNS Neurosci Ther. 2024;30(3): e14664.
55. Huang H, Bhat A, Woodnutt G, Lappe R. Targeting the ANGPT-TIE2 pathway in malignancy. Nat Rev Cancer. 2010;10(8):575–85.
56. Zhang K, Wang J, Zhu Y, Liu X, Li J, Shi Z, et al. Identification of Hub Genes Associated With the Development of Stomach Adenocarcinoma by Integrated Bioinformatics Analysis. Front Oncol. 2022;12: 844990.
57. Li J, Gao S. HOXB5-activated ANGPT2 promotes the proliferation, migration, invasion and angiogenic effect of esophageal cancer cells via activating ERK/AKT signaling pathway. Exp Ther Med. 2022;24(3):585.

Publisher's Note