ANALYTICAL AND CLINICAL VALIDATION OF A ctDNA BASED ASSAY FOR MULTI-CANCER EARLY DETECTION

Luu Hong Dang Nguyen[1#], Thi Hue Hanh Nguyen[1#], Van Hoi Le[2], Vinh Quang Bui[3], Lan Hieu Nguyen[4], Nhu Hiep Pham[5], Thanh Hai Phan[6], Huu Thinh Nguyen[7], Van Song Tran[8], Chi Viet Bui[9], Van Kha Vo[10], Pham Thanh Nhan Nguyen[11], Ha Huu Phuoc Dang[12], Van Dung Pham[13], Van Thinh Cao[14], Ngoc Minh Phan[1], Ba Linh Tieu[1], Giang Thi Huong Nguyen[1], Dac Ho Vo[1], Trung Hieu Tran[1], Thanh Dat Nguyen[1], Van Thien Chi Nguyen[1], Trong Hieu Nguyen[1], Vu Uyen Tran[1], Minh Phong Le[1], Thi Minh Thu Tran[1], Minh Nguyen Nguyen[1], Thi Tuong Vi Van[1], Anh Nhu Nguyen[1], Thi Thanh Nguyen[1], Huu Tam Phuc Nguyen[1], Dinh Yen An Truong[1], Y-Thanh Lu[1], Chi Thuy Tien Cao[1], Van Tung Nguyen[2], Thi Le Quyen Le[3], Thi Lan-Anh Luong[4], Thi Kim Phuong Doan[4], Thi Trang Dao[4], Canh Duy Phan[15], Thanh Xuan Nguyen[15], Nguyen Tuong Pham[15], Bao Toan Nguyen[6], Thi Thu Thuy Pham[6], Huu Linh Le[6], Cong Thanh Truong[6], Thanh Xuan Jasmine[6], Minh Chi Le[7], Van Bau Phan[8], Quang Binh Truong[7], Thi Huong Ly Tran[10], Minh Thien Huynh[10], Tu Quy Tran[11], Si Tuan Nguyen[13], Vu Tran[13], Van Khanh Tran[14], Huu Nguyen Nguyen[1], Duy Sinh Nguyen[1], Thi Van Phan[1], Thi Thanh-Thuy Do[16], Dinh Kiet Truong[16], Hung Sang Tang[1], Hoa Giang[1], Hoai-Nghia Nguyen[1], Minh-Duy Phan[1 *], Le Son Tran[1*]

[1]Gene Solutions, Ho Chi Minh, Vietnam
[2]National Cancer Hospital, Hanoi, Vietnam
[3]Hanoi Oncology Hospital,Hanoi, Vietnam
[4]Hanoi Medical University, Hanoi, Vietnam
[5]Hue Central hospital, Hue, Vietnam
[6]Medic Medical Center, Ho Chi Minh, Vietnam
[7]University Medical Center HCM, Ho Chi Minh Vietnam
[8]People's Hospital 115, Ho Chi Minh, Vietnam
[9]Xuyen A General Hospital, Ho Chi Minh, Vietnam
[10]Cantho Oncology Hospital, Can Tho, Vietnam

[11]Danang Oncology Hospital, Da Nang, Vietnam
[12]Dongnai General Hospital, Dong Nai, Vietnam
[13]Thong Nhat Dongnai General Hospital, Dong Nai, Vietnam
[14]Le Van Thinh Hospital, Ho Chi Minh, Vietnam
[15]Hue Central hospital, Hue, Vietnam
[16]Medical Genetics Institute, Ho Chi Minh, Vietnam

[#] Luu Hong Dang Nguyen and Thi Hue Hanh Nguyen contributed equally to this study.

*Correspondence: leson1808@gmail.com, pmduy@yahoo.com

Key words: liquid biopsy, multimodal analysis, cell-free DNA, circulating tumor DNA, multicancer early detection, tissue of origin, analytical validation, and clinical validation.

ABSTRACT

The development of multi-cancer early detection (MCED) through a single blood test has emerged as a promising method for improving the efficiency of early cancer detection and benefiting population health. However, limits to the use of MCED tests in clinical practice - arising from the lack of analytical validation and clinical evidence for their utility in diverse populations, have prevented their application. To address these challenges, we conducted a comprehensive analytical and clinical validation for a MCED test, SPOT-MAS (Screening for the Presence Of Tumor by DNA Methylation And Size) in this study. The analytical validation was reported to establish the clinical limit of detection, reproducibility of test results and assess the impact of potential interferents on test performance. Our assay demonstrated its robustness in identifying cancer patients with a limit of detection of 0.049 across five tumor cancer types. In the reproducibility study, results were consistent in performance for both intra- and inter-batch analysis. Moreover, our test remained robust at hemoglobin contamination of 500 mg/dl and genomic DNA contamination of 100%. To validate the performance of SPOT-MAS test in clinical setting, we launched a multi-center prospective trial of 10,027 asymptomatic participants in Vietnam, named K-DETEK. Our assay obtained a positive predictive value of 58.14% with 84.00% accuracy in predicting tumor location and a negative predictive value of 99.92%. To our knowledge, this is the first and largest prospective validation study in Asia supporting the utility of SPOT-MAS as a multi-cancer blood test for early detection in a low- and middle-income country, where a nationwide cancer screening program is urgently needed but currently not available.

INTRODUCTION

Cancer is the second leading cause of death and continues to have a significant impact on global mortality rates. The prevalence of cancer and its associated mortality has put a strain on healthcare systems worldwide (*1*). Detection in advanced stages of disease has also increased the burden of cancer (*2*). In the effort to combat this, current cancer screening methods, such as those recommended by the United States Preventive Services Task, have shown promise in increasing the overall survival rate, improving treatment efficiency, and reducing long-term medical costs. However, some conventional screening methods, particularly colonoscopy or cervical cytology tests, are invasive, resulting in low accessibility. Moreover, current screening tests only test for a single cancer type, leading to a high accumulative false positive rate and consequent overdiagnosis and overtreatment when performed sequentially. One of the most promising advancements in cancer detection is the development of a non-invasive multi-cancer early detection (MCED) test (*3*). In MCED, the blood-based liquid biopsy (LB) approaches have the potential to revolutionize cancer screening by enabling early detection of multiple types of cancer through a simple blood draw (*4*). LB assays detect specific cancer-related biomolecules including circulating tumour cells (CTC), circulating tumor DNA (ctDNA), circulating free RNA (cfRNA) and exosomes (*5-8*). Of these, ctDNA released into the circulation when tumor cells undergo apoptotic and necrotic cell death processes has been extensively studied due to its tissue- and cancer-type specificity (*5, 6, 9-11*). Recently, the landscape of MCED tests based on detecting methylation changes in cfDNA has evolved. The OverC test (Burning Rock) or Galleri test (Grail) have shown to detect multiple cancer types simultaneously with high performance by interrogating methylation changes in cfDNA (*12, 13*). The Galleri test, has been clinically validated in asymptomatic (PATHFINDER study) and symptomatic population (SYMPLIFY

study). In recent publications, they showed promising data for the test performance with a PPV of 43.1% and the tumor of origin (TOO) prediction with an accuracy of 88% in the asymptomatic screening population (*12*). Moreover, for individuals with malignancy-related symptoms, the data from SYMPLIFY trial in England and Wales demonstrated a particularly high PPV of 75.5% and NPV of 97.6% (*13*).

Despite promising results, current MCED methods demonstrated low sensitivity for detecting certain cancers (e.g., breast cancer) and early-stage tumors owing to low amount and high heterogeneity of ctDNA. To improve the detection sensitivity of ctDNA, current screening methods tend to use high-depth sequencing, making it economically impractical for population-wide screening. To address these limitations, we recently developed a multimodal method, known as Screening for the Presence Of Tumor by Methylation And Size (SPOT-MAS), to simultaneously detect five common types of cancer, including liver, breast, colorectal, gastric, and lung cancer, and predict the cancer signal origin (*14*, *15*). By integrating cost-effective shallow sequencing and advanced machine learning, SPOT-MAS were trained and validated on a large cohort of 2,288 participants, including 738 nonmetastatic patients and 1,550 healthy controls, with a sensitivity of 72.4% at 97.0% specificity and an accuracy of 70% for tumor-of-origin (TOO) prediction. We further assessed the performance of SPOT-MAS in an interim 6-month study, named K-DETEK, which involved 2,795 participants at 14 clinical centers in Vietnam (*16*). In this study, we present the analytical validation of SPOT-MAS assay to determine the sensitivity and investigate the impact of the technical factors on the robustness of our assay, which have not been reported in our interim report (**Figure 1**). Moreover, we comprehensively evaluated the clinical performance of SPOT-MAS on 8,973 participants who had completed the 12-month follow-up period to prove its utility in the early detection of

multiple type cancer types (**Figure 1**). The clinical applicability of SPOT-MAS is reported by quantitatively presenting positive predictive value (PPV), negative predictive value (NPV) and accuracy of TOO prediction.

METHODS

Analytical validation

Three investigations were performed to validate this study, including test limit of detection (LOD), potential interferents and reproducibility (**Figure 1**).

Analytical sample preparation

For analytical analysis, samples were collected from healthy individuals and diagnostic-confirmed for each type of cancer from our previous case-control study (*17*). Healthy standard participants were confirmed cancer-free within three years. For each cancer type among breast, colorectal, gastric, liver, and lung, standard samples were confirmed from patients with cancer diagnosis in their early stages by pathology, and imaging results. To obtain sufficient materials for analytical validation, we generated pooled cfDNA or pooled plasma samples. Specifically, pooled cfDNA samples were generated by mixing cfDNA isolated from different healthy subjects (n = 96) or cancer patients (n = 169) diagnosed with the same cancer type and used for LOD determination (**Table S1**). Pooled plasma samples were generated by mixing 1 ml of plasma isolated from healthy subjects (n = 25) or patients (n = 50) diagnosed with the same cancer type and used for inference and test reproducibility assessment (**Table S1**).

Tumor fraction estimation and establishing clinical limit of detection

In this study, 65 pooled plasma samples from healthy individuals and 54 pooled plasma samples from cancer patients, including breast cancer (n= 13), colorectal cancer (n= 12), gastric cancer

(n= 6), liver cancer (n= 10) and lung cancer (n= 13), were subjected to SPOT-MAS assay. Then, the tumor fraction (TF) of cancer cfDNA samples were determined by ichorCNA tool which predicts segments of somatic copy number alterations (CNAs) and estimate TF while accounting for subclonality and tumor ploidy (*18*). Then, the cancer cfDNA samples with known TF were spiked into the healthy cfDNA samples with various levels of 0.5, 1, 5, 15, 25, 50 and 100% of the neat samples with estimated TF. Six to twelve replicates of each TF level for each cancer type were investigated. To define LOD of SPOT-MAS for each type of cancer, we defined as the TF levels at which the probability of detecting a cancer signal was at least 50% while maintaining at least 98% specificity as previously described by Jamshidi et al. (*19*).

Potential interferents

To determine the effect of potential substances that could interfere with the performance of SPOT-MAS assay, 10 pooled cancer and 5 healthy-control plasma samples were spiked with genomic DNA or hemoglobin. Each pooled plasma sample was spiked with genomic DNA (0, 50, 100, 150, 200%, equivalent to 0, 50-105, 100-215, 150-320, 200-430 ng) or hemoglobin (0, 100, 500, 1000 and 2000 mg/dL).

Reproducibility study

To determine the ability to reproduce and variability of SPOT-MAS assay, a total of 51 samples including 27 cancer and 24 healthy-control. The concordance rate was analyzed within runs (intra-batch) and between runs (inter-batch) with different lots of reagents and operators. For the analysis of intra-batch variations, 24 healthy controls and 27 cancer samples were performed in triplicate (denoted by A, B and C). For the analysis of inter-batch, a total of 24 healthy samples and 27 cancer samples were divided into the three independent batches (denoted by 1, 2, and 3).

Samples were subjected to SPOT-MAS assay and the consistency was evaluated based on the prediction results of SPOT-MAS.

Clinical validation

Study design

K-DETEK is a prospective study for participants attending to annual health check-ups across 13 hospitals and one research institute in Vietnam from April 2022 to October 2022. The trial was registered with the U.S. National Institutes of Health (ClinicalTrials.gov identifier: NCT05227261). The institutional ethics and scientific committee of the University of Medicine and Pharmacy, Ho Chi Minh City, Vietnam, reviewed and approved this study (approval number: 192/HĐĐĐ-ĐHYD). All participants were provided written informed consent for prior to participating in study-related activities and sample collection for K-DETEK study. Participants were eligible for K-DETEK study if they were aged 40 years or older, willing to return for required follow-up visits at 6 months and 12 months, had neither clinical suspicion of cancer nor history of confirmed cancer, had no history of blood transfusion or bone marrow transplantation in the past 3 months and had no clinical manifestations of pregnant.

Workflow

Plasma cfDNA was extracted from 10 ml blood. The median time from blood collection to plasma isolation was 2 days, range from 0 to 5 days. The plasma collection and storage, the cfDNA extraction and storage, and the cfDNA sodium bisulfite treatment were performed as described earlier. Briefly, the multiple features of isolated cfDNA, including 450 target regions' methylation, genome-wide methylation profiles (global methylation density of 2734 1Mb-bins on 22 chromosomes), fragment length, DNA copy number of 588 5Mb-bins on 22 chromosomes

and end motif, were simultaneously analyzed by SPOT-MAS assay. The SPOT-MAS procedure was previously described by our group (cite). Using machine learning algorithms, the model returned the probability scores of ctDNA detection and the tumor of origin (TOO) for those with ctDNA signal detected (**Figure 1**).

Outcomes

The SPOT-MAS test results were returned to the study participants within 30 days from the blood collection. SPOT-MAS provides two test results, including "ctDNA signal detected" (positive) or "ctDNA signal not detected" (negative) and up to two prediction results of TOO.

The participants with a detected ctDNA signal were consulted by physicians to undertake the diagnostic imaging tests according to the TOO probability values across five types of cancer: breast, liver, lung, colorectal and gastric. TOO signals that are not covered by SPOT-MAS test were reported as "other cancer". Cancer diagnoses were reported according to The National Comprehensive Cancer Network (NCCN) for cancer screening. The diagnosis tests were listed in **Table S2**. Participants with "other cancer" status were advised to check whole body CT scans for cancer screening. Participants with no cancer signal detected were advised about their lack of risk for the five cancer types covered by the SPOT-MAS test. All participants were followed up after 6 months and 12 months to obtain information on possible cancer diagnosis.

The performance of SPOT-MAS assay was determined, including true-positive rate, false-positive rate, positive predictive value (PPV) and negative predictive value (NPV) for cancer signal detection (%). The overall prediction accuracy (%) of TOO was also assessed.

Participant demographic and statistical analysis

The demographic information of all participants was listed in **Table 1**. Participants in high-risk group included factors such as smoking, drinking, hepatitis B/C or having First-Degree Relatives (FDR) diagnosed two cancer types at the age of younger than 45 or identified mutant carries. The remaining participants were classified in moderate risk group.

All statistical analyses were carried out using R (4.1.0) with standard data analysis packages and the ggplot2 package for visualization.

RESULTS

Determination of limit-of-detection of SPOT-MAS assay

Limit of detection (LOD) is a critical parameter to assess the analytical sensitivity of an assay. To evaluate the LOD value of SPOT-MAS test, we first generated cancer standard samples for five different types of cancer, including breast, colorectal, liver, lung and gastric cancer by pooling cfDNA samples of patients diagnosed with the same cancer type (**Figure 2A**). Subsequently, we determined tumor fraction (TF) of these pooled cfDNA samples by using the ichorCNA tool that was previously developed by Viktor et al.(*18*). The icorCNA tool was originally developed to quantify TF in low-pass non-bisulfite genome-wide sequencing reads. To demonstrate the feasibility of using ichorCNA for our bisulfite-converted cfDNA, we randomly selected cfDNA from 6 cancer and 4 healthy plasma samples to perform a comparison of TF levels between bisulfite and non-bisulfite sequencing results. We observed an excellent correlation between tumor fraction of bisulfite-treated and non-bisulfite-treated fragments calculated by ichorCNA for all 10 samples (Pearson correlation, $R^2 = 0.99$, $p= 0.17.10^{-11}$, **Figure S1**), indicating that the ichorCNA tool can be applied to bisulfite-converted sequencing cfDNA samples. After quantifying TF of our pooled cancer cfDNA samples, we spiked the cancer cfDNA with known TF into the healthy cfDNA samples with different levels of 0.5, 1, 5, 15, 25, 50 and 100% of the

total estimated TFs. As a result, we obtained seven spike-in cfDNA samples with different TF levels and 6-15 replicates for each cancer type (**Figure 2B**). We observed positive linear correlations between TF levels and spike-in concentrations, starting from 5% for colorectal and liver cancer, from 25% for lung and gastric cancer. In contrast, the neat cfDNA samples (100%) from breast cancer had the lowest TF among the five investigated cancer types (median 0.041) and displayed no linear correlation across different spike-in standard samples.

To determine LOD value, we first performed the SPOT-MAS test on 378 replicates with estimated TFs from the seven standard samples of each cancer type and 65 cfDNA samples from healthy individuals. We found that 65/65 (100%) were called correctly as non-cancer by SPOT-MAS test, 135/378 (35.7%) were detected as cancer samples including 43/84 (51.19%) colorectal, 43/70 (61.43%) liver, 31/91 (34.07%) lung, 6/42 (14.29%) gastric and 12/91 (13.18%) breast cancer (**Figure 2B**). We next determined the LOD value for each cancer type by defining the TF at which at least 50% of cancer samples were correctly detected by SPOT-MAS while maintaining a 98% specificity (**Figure 2C**). The overall LOD for detecting all cancer samples was 0.049 (95% CI 0.044-0.057, **Figure 2D**). The lowest LOD was determined for lung cancer (0.041, 95% CI 0.033-0.068, **Figure 2D**), followed by liver (0.049, 95% CI 0.043-0.059, **Figure 2D**), colorectal (0.051, 95% CI 0.045-0.060, **Figure 2D**), breast (0.052, 95% CI 0.041-0.18, **Figure 2D**), and gastric (0.081, 95% CI 0.055-0.12, **Figure 2D**). Our data suggested that SPOT-MAS could detect at least 50% of cancer samples at a specificity of 98% if the samples have ichorCNA estimated TF of greater than 0.049.

Evaluation the impact of potential interferents on the performance of SPOT-MAS

Our previous study showed that the quality of collected blood samples, particularly the hemolysis rate, is an important determinant of DNA library efficacy (*20*). It has been well

established that blood hemolysis causes genomic DNA and hemoglobin contamination which could interfere the performance of liquid biopsy tests (*20*). To assess the impact of these two potential interferents, we generated 15 pooled plasma samples, including cancer (N=10) and healthy-control (N=5). We next spiked into these pooled plasma samples with five different concentrations of genomic DNA or hemoglobin (**Figure 3A**). All healthy control samples were correctly detected across different levels of spiked-in gDNA (**Figure 3B**) or hemoglobin (**Figure 3C**). By contrast, the false negative rates were increased from 13.3% to 26.7% when the samples were contaminated with a gDNA concentration of higher than 100% (**Figure 3D**). Likewise, the false negative rates were increased by 6.7% when the levels of hemoglobin contamination were greater than 500 mg/dL (**Figure 3E**). Thus, our findings indicated that the gDNA and hemoglobin contamination, possibly in samples with high levels of hemolysis could affect the accuracy of SPOT-MAS performance.

Evaluation of the reproducibility of SPOT-MAS

To examine the reproducibility of SPOT-MAS, we used a total of 51 pooled plasma samples (27 cancers and 24 healthy controls) and divided them into three different batches (**Figure 4A**). For the analysis of intra-batch variations, 24/24 (100%) healthy samples 26/27 (96.30%) cancer samples were correctly detected by SPOT-MAS test across three different replicates (A, B and C), except for one breast cancer sample was incorrectly detected as healthy sample in batch B (**Figure 4B**). For the evaluation of inter-batch variations, we observed high levels of consistency, with 50/51 (98.04%) samples being correctly detected, except for one breast sample in batch 3 which was incorrectly detected as healthy sample (**Figure 4C**). These findings showed a highly consistent performance of SPOT-MAS assay within runs or between runs.

Evaluation of clinical performance of SPOT-MAS test

A total of 10,027 participants (**Table S3**) were enrolled in the K-DETEK study from April 2022 to September 2023 from 13 hospitals and 1 research institute in Vietnam (**Figure 5**). Of those, 9,024 (90.00%) eligible participants had blood samples collected and received SPOT-MAS test results while 1,003 participants (10.00%) were excluded from the study. The reasons for exclusion included a diagnosis of cancer (n = 1), pregnancy (n = 1), high levels of blood hemolysis (n = 31) or providing incorrect follow up details or being lost contact at 6-month follow-up period (n = 970) (**Figure 5**).

To date, 6-month follow-up surveys have been completed for all participants and 8,793 of 9,024 (97.44%) participants have completed the 12-month follow-up (**Figure 5**). The clinical characteristics of those participants are summarized in **Table 1**, with a higher percentage of females than males (54.59% versus 45.41%, **Table 1**) and a median age of 50□years, ranging from 40 to 94 years. High-risk individuals who harbor the following risk factors including smoking, drinking, hepatitis B/C, FDR (First-Degree Relatives diagnosed with two cancer types), accounted for 26.23% (n□=□2,306) of all participants. The remaining 73.77% (n□=□6,487) individuals were considered moderate risk (Table 1).

The vast majority of participants (8,750/8,793; 99.51%) with “ctDNA signal not detected” and 8,743 (99.92%) of them were confirmed to be cancer-free at 12 months after enrollment, indicating a NPV of 99.92% (**Table 2**). Seven (0.08%) cases were found to develop cancer during the 12-month follow-up (**Table S4**), with 2 cases developing metastatic lung cancer (n=2, patient K00044 and K00049) , 2 cases with locally advanced colorectal cancer (n= 2, patient K00048 and K00050), 3 cases developing localized cancer including colon (patient K00045), lung (patient K00046) and gastric cancer (patient K00047).

We detected 43 cases (0.49%) with “ctDNA signal detected” who were refered to undertake on-site standard of care (SOC) imaging tests to confirm the presence of tumors according to our consultation protocol (**Table S2**). All of them agreed to undertake diagnostic tests for cancer types based on the prediction of TOO provided in the SPOT-MAS test report. Among the 43 participants with confirmed diagnostic results, 29 had imaging results with lesions suspected of malignancy and were advised to perform tissue biopsies. Of those cases, 25 were subsequently shown to have cancer or pre-cancerous lesions detected in different primary organs, suggesting a PPV of 58.14% (**Table 2**). Among the 25 true positive cases, 21 developed cancer in the organs matched with either the first or second cancer type predicted by SPOT-MAS, suggesting an overall accuracy of 84.00% for TOO (**Figure 6**). Specifically, 4 cases had metastatic colorectum, prostate, and cholangiocarcinoma (patients K00002, K00003, K00009, and K00013, **Figure 6**), 3 cases were diagnosed with locally advanced cancer at gastric and lung (patients K00006, K00007, and K00008, **Figure 6**), and 6 cases were diagnosed with localized cancer in different primary organs, including endometrium, breast, liver, lymphoma and colon (patients K00001, K00004, K00005, K00010, K00011, and K00012, **Figure 6**). The remaining 12 true positive cases were shown to have premalignant lesions belonging to the dysplastic or metaplastic group (patients from K00014 to K00025, **Figure 6**). Four cases (patient K00001, K00003, K00011, and K00013, **Figure 6**) were found to have other tumor types that were not covered by SPOT-MAS (**Table S4)** including 1 endometrial cancer (patient K00001), 1 prostate cancer (patient K00003), 1 Jejunal mesenteric lymphoma (patient K00011), and 1 cholangiocarcinoma cancer (patient K00013). The remaining 18 cases (0.20%, **Table 2**) with diagnosis diagnostic results unable to confirm the presence of malignant or precancerous tumors were regarded as false positive cases. However, we noted that 77.78% (14/18) of them developed benign lesions and 10 showed benign

lesions such hyperplastic polyps (patient K00042), BI-RADS 4 breast fibroadenoma (patient K00026), in the organs matched with the TOO prediction by SPOT-MAS (**Figure 6**). We observed that the median time from receipt of positive SPOT-MAS results to final diagnosis confirmation was 19 days ranging from 1 to 187 days for all 40 cases, with false positive group having a shorter median time (13 days) as compared with true positive group who need 22 days to achieve diagnostic resolution (**Figure 6**).

Since previous studies have reported that the performance of a MCED could be dependent on the risk of target populations, we next examined such association in our cohort. We did not observe any noticeable difference in NPV across diverse groups of participants (NPV > 99.8%, **Figure 7**). By contrast, the PPV increased from 56.25% in moderate risk participants to 63.64% in high-risk participants. Moreover, we observed higher PPV in the group over 50 years old as compared to the younger group < 50 years old (61.11% versus 42.86%, **Figure 7**). For gender, the PPV was higher in male participants than female participants (71.43% versus 51.72%, **Figure 7**).

Overall, our data shows that SPOT-MAS test could detect a variety of cancer types either at pre-cancerous or early-stages with a fixed PPV of 58.14% and a NPV of 99.92%. Apart from detecting cancer signal, SPOT-MAS was able to identify the tissue localization of cancer signal with an accuracy of 84.00%, that can assist clinicians in directing patients to appropriate diagnostic tests.

Discussion

The paradigm of cancer diagnosis is undergoing a significant shift with the development of MCED tests. MCED in a single blood draw test is key to the successful treatment and improve survival outcomes for cancer patients. To ensure the effectiveness and reliability of MCED tests in clinical practice, thorough validation is crucial. Here, we conducted the analytical validation to

determine the clinical LOD and evaluate the impact of various technical factors on the robustness and accuracy of our MCED test, named SPOT-MAS. Moreover, we presented the performance of SPOT-MAS test on a large-scale cohort of 8,793 asymptomatic participants with 12-month follow-ups.

To determine the LOD for SPOT-MAS test, we applied ichorCNA software to estimate TF of pooled cfDNA samples isolated patients with five types of cancer. We next generated a series of cfDNA standards with different TF levels by spiking cancer cfDNA into healthy cfDNA. Of these cancer types, SPOT-MAS demonstrated the lowest LOD value for detecting lung cancer (0.041) while it showed the highest LOD value for detecting breast (median LOD 0.052) gastric cancer (median LOD 0.081). The low sensitivity for detecting ctDNA signal in breast and gastric samples was consistent with our previous study that validated the performance of SPOT-MAS in these two types of samples (*17*). Our findings are in line with previous studies showing low levels of ctDNA shedding by early-stage gastric and breast tumors (*4*, *15*, *21*).

Our previous findings showed the relationship between high hemolysis rate and its effect on isolated cfDNA and sequencing quality (*20*). Therefore, we evaluated their potential interferents on our assay performance in this study. Although SPOT-MAS test remained robust at hemoglobin contamination of 500 mg/dl or gDNA contamination of 100%, its performance was reduced beyond these threshold values (**Figure 3B**). Our findings are in good agreement with previous studies showing that gDNA displayed methylation and fragment profile distinct from ctDNA (*22*-*24*) thereby possibly introducing biased signals to our test. Therefore, the quality of blood samples should be meticulously controlled to ensure the high quality of the test.

Although genome-wide low-pass sequencing was incorporated into SPOT-MAS workflow to ensure cost effectiveness, this approach might introduce batch effects. Our assay still observed

high consistency of performance for both intra- and inter-batch analysis (**Figure 4**). This could be attributed to the nature of our assay which interrogated multi-features instead of a single feature of ctDNA to achieve a final ctDNA probability score, thus alleviating the variability (*17*). Furthermore, except blood collection, all the important steps including library preparation and sequencing were conducted at the same central laboratory.

The SPOT-MAS test was constructed to detect the top five common types of cancers, that account for more than 60% of cancer cases in Vietnamese population with a cancer incidence of 0.15% (*25*). Our study revealed a high positive rate of 0.49%, 43 of 8,793 eligible participants had a cancer signal detection. The selection criteria of our study which targeted moderate- and high-risk participants, possibly contributed to the high positive rate.

In this study, the majority of participants had no ctDNA signal detection, with a NPV of 99.92%. There were 7 participants developing cancer within the 12-month follow-up period. We found that 2 out of 7 negative cases (patient K00044 and K00050) developed metastatic cancer. This misidentification could be due to the variations in methylation or fragment profiles between late-stage and early-stage tumors. Metastatic tumors display methylation signatures different from primary early-stage tumors and may not be detected by our algorithms which learned those signatures from samples with rigor selection criteria for non-metastatic cancer (stage I-II and IIIA) in our training dataset (*26*). Moreover, 30% - 60% of patients with metastatic cancer have cancer of unknown primary sites which may be out of the scope of the SPOT-MAS test (*27*). Moreover, we performed TF analysis by ichorCNA analysis for the 7 false negative cases and none of them displayed TF above the overall LOD of our test for detecting 5 cancer types (**Table S6 and Figure S2**). This finding suggested that they might have low amounts of ctDNA in plasma which could not be captured by SPOT-MAS test.

Our findings from the K-DETEK study showed that SPOT-MAS test achieved a PPV of 58.14%, indicating its effectiveness in correctly identifying over half of participants with a cancer diagnosis. For the 25 true positive cases, our analysis observed an accuracy of 84.00% for predicting TOO, thus highlighting the potential utility of SPOT-MAS test for clinicians in guiding their standard-of-care decisions and minimizing the risk of overdiagnosis. Moreover, our stratification analysis showed higher PPV in older age, male patients or those belonging to high-risk group. This finding indicates the importance of considering clinical characteristics and demographic of screening populations, when evaluating test performance and outcomes. Among true positive cases, 12 were found to have precancerous polyps in their stomach or colon. Researchers have indicated that these polyps, including metaplasia and dysplasia, carry a risk of malignant cancer development ranging from 6% to 35% (*28, 29*). This underscores the critical role of SPOT-MAS in preventing the progression to malignant cancer, thereby possibly improving clinical outcomes for cancer patients. Moreover, our test only took 19 days to achieve diagnostic resolution. Diagnostic resolution in a short period of time could significantly alleviate patients' nervousness, expedite the necessary interventions for those diagnosed with precancerous conditions.

There were 18 cases who initially had ctDNA signal detected, but their diagnostic tests could not confirm the presence of tumor. It is possible that they might have early-stage tumors but with small size that could not be captured by SOC imaging tests. Alternatively, they might have tumor types that did not match with the prediction of SPOT-MAS or they might have special pathological condition at the time of taking SPOT-MAS test, resulting in generating ‘pseudo signal’ (*30, 31*).

In comparison with other clinical studies, it is essential to consider the variations in population risk and their implications on test performance (*12, 13, 32*). Our K-DETEK study focused on moderate- and high-risk participants with a high chance of developing cancer, resulting in a higher PPV than PATHFINDER, which evaluated the MCED test in adults over than 50 years or older without signs or symptoms of cancer. Moreover, our study observed a lower PPV compared to SYMPLIFY, which concentrated on symptomatic patients with median age of 61.9 years. Importantly, our study detected a wide range of solid tumors and precancerous lesions, whereas the majority of cancer patients identified in PATHFINDER had hematologic cancers (48.57 %). The multimodal approach of SPOT-MAS offers a comprehensive analysis through the integration of methylation, fragment length profile, DNA copy number aberration and end motif in a single library reaction. This unique approach could explain the differences observed in the comparison to the GALLERI test used in the PATHFINDER study, which primarily focused on methylation makers (*17, 33*).

Our study has several limitations. First, the determination of limit of detection by ichorCNA tool might not provide an accurate reflection of the true LOD of SPOT-MAS test, particularly those cancer types with low CNA signals because our assay identifies cancer signals based on multiple ctDNA signatures. Second, a small proportion of participants (231/9024, 2.56%) have not completed the 12-month follow-up point, which might have an impact on the precise estimation of NPV. Third, our assay has low sensitivity for detecting certain cancer types with low shedding tumors, such as gastric and breast, or metastatic tumors, resulting in false negative cases. Moreover, SPOT-MAS only focuses on five common cancer types, potentially missing rarer cancer types. To further enhance the sensitivity of early cancer detection and provide the broader spectrum of cancer types, our future studies will explore the multi-omics approach that combine

different makers, such as cfDNA, cfRNA and circulating tumor cells (*34, 35*). Fourth, 18 false positive cases in this study could lead to unnecessary diagnosis procedure and 14 of them experienced invasive endoscopy tests. However, SPOT-MAS has demonstrated promising results in maintaining a fixed PPV of 58.14% for 5 cancer types, thereby avoiding an accumulated false positive rate when compared to single cancer type SOC tests. Finally, it remains unclear whether a MECD test such as SPOT-MAS could enhance survival benefits, future randomized trial studies are needed to address this question. These studies will provide valuable insights into the effectiveness of the test and its impact on patient outcomes.

In conclusion, our study demonstrated SPOT-MAS as a promising multi-cancer blood test for early cancer detection. Through comprehensive analytical validation, we have established the robustness of SPOT-MAS test, demonstrating a limit of detection of 0.049. Importantly, the data from K-DETEK study, the largest prospective trial in Asia, highlights the potential utility of SPOT-MAS in a low- and middle-income countries. Together, our study provides convincing evidence to support the feasibility of implementing SPOT-MAS to benefit the population's health by effectively identifying cancer patients in the asymptomatic population.

Declarations

Ethics approval and consent to participate:

This study was approved by the Ethics Committee of the Medic Medical Center, University of Medicine and Pharmacy and Medical Genetics Institute, Ho Chi Minh city, Vietnam. Written informed consent was obtained from each participant in accordance with the Declaration of Helsinki.

Consent for publication:

Not applicable.

Availability of data and materials:

Sequencing data will be deposited in a public portal database (NCBI SRA) upon acceptance and are available on request from the corresponding author, LST. The data are not publicly available due to ethical restrictions.

Competing interests:

The authors declare no conflict of interest.

Funding:

The study was funded by Gene Solutions

Disclosure statement:

The authors including LST, HNN, HG, MDP, HHN and DSN hold equity in Gene Solutions. The funder Gene Solutions provided support in the form of salaries for authors are inventors on the patent application (USPTO 17930705). We also confirm that this does not alter our adherence to Cancer Investigation policies on sharing data and materials.

Author contribution:

Formal analysis: Luu Hong Dang Nguyen, Thi Hue Hanh Nguyen, Ngoc Minh Phan, Ba Linh Tieu, Dac Ho Vo, Trung Hieu Tran, Thanh Dat Nguyen, Van Thien Chi Nguyen, Y-Thanh Lu, Trong Hieu Nguyen, Vu Uyen Tran, Minh Phong Le, Thi Minh Thu Tran, Minh Nguyen Nguyen, Thi Tuong Vi Van, Anh Nhu Nguyen, Thi Thanh Nguyen, Huu Tam Phuc Nguyen, Dinh Yen An Truong, Chi Thuy Tien Cao

Patient consultancy and screening: Luu Hong Dang Nguyen, Thi Hue Hanh Nguyen, Van Hoi Le, Vinh Quang Bui, Lan Hieu Nguyen, Nhu Hiep Pham, Thanh Hai Phan, Huu Thinh Nguyen, Van Song Tran, Chi Viet Bui, Van Kha Vo, Pham Thanh Nhan Nguyen, Ha Huu Phuoc Dang, Van Dung Pham, Van Thinh Cao, Ba Linh Tieu, Thi Thanh Nguyen, Huu Tam Phuc Nguyen, Dinh Yen An Truong, Chi Thuy Tien Cao, Van Tung Nguyen, Thi Le Quyen Le, Thi Lan-Anh Luong, Thi Kim Phuong Doan, Thi Trang Dao, Canh Duy Phan, Thanh Xuan Nguyen, Nguyen Tuong Pham, Bao Toan Nguyen, Thi Thu Thuy Pham, Huu Linh Le, Cong Thanh Truong[6], Thanh Xuan Jasmine, Minh Chi Le, Van Bau Phan, Quang Binh Truong, Thi Huong Ly Tran, Minh Thien Huynh, Tu Quy Tran, Si Tuan Nguyen, Vu Tran, Van Khanh Tran, Huu Nguyen Nguyen, Duy Sinh Nguyen, Thi Van Phan, Thi Thanh-Thuy Do, Dinh Kiet Truong, Hung Sang Tang

Data Curation: Luu Hong Dang Nguyen, Thi Hue Hanh Nguyen

Methodology: Duy Sinh Nguyen, Thi Van Phan, Thi Thanh-Thuy Do, Dinh Kiet Truong, Hung Sang Tang, Minh-Duy Phan, Hoa Giang, Hoai-Nghia Nguyen, Le Son Tran

Conceptualization: Vinh Quang Bui, Lan Hieu Nguyen, Nhu Hiep Pham, Thanh Hai Phan, Huu Thinh Nguyen, Van Song Tran, Chi Viet Bui, Van Kha Vo, Pham Thanh Nhan Nguyen, Ha Huu Phuoc Dang, Van Dung Pham, Van Thinh Cao, Van Tung Nguyen, Thi Le Quyen Le, Thi Lan-Anh Luong, Thi Kim Phuong Doan, Thi Trang Dao, Canh Duy Phan, Thanh Xuan Nguyen, Nguyen Tuong Pham, Bao Toan Nguyen, Thi Thu Thuy Pham, Huu Linh Le, Cong Thanh Truong[6], Thanh Xuan Jasmine, Minh Chi Le, Van Bau Phan, Quang Binh Truong, Thi Huong Ly Tran, Minh Thien Huynh, Tu Quy Tran, Si Tuan Nguyen, Vu Tran, Van Khanh Tran, Huu Nguyen Nguyen, Duy Sinh Nguyen, Thi Van Phan, Thi Thanh-Thuy Do, Dinh Kiet Truong, Hung Sang Tang, Minh-Duy Phan, Hoa Giang, Hoai-Nghia Nguyen, Le Son Tran

Writing-original draft: Luu Hong Dang Nguyen, Thi Hue Hanh Nguyen, Giang Thi Huong Nguyen, Le Son Tran

Writing-Review and Editing: Giang Thi Huong Nguyen, Minh-Duy Phan, Hoa Giang, Hoai-Nghia Nguyen, Le Son Tran

Acknowledgments:

We thank all participants who agreed to take part in this study, and all the clinics and hospitals who assisted in patient consultation and sample collection.

References

1. T. Pham *et al.*, Cancers in Vietnam-Burden and Control Efforts: A Narrative Scoping Review. *Cancer Control* **26**, 1073274819863802 (2019).
2. N. Hawkes, Cancer survival data emphasise importance of early diagnosis. *BMJ* **364**, l408 (2019).
3. Z. Kakushadze, R. Raghubanshi, W. Yu, Estimating Cost Savings from Early Cancer Diagnosis. *Data*. 2017 (10.3390/data2030030).
4. M. C. Liu, G. R. Oxnard, E. A. Klein, C. Swanton, M. V. Seiden, Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol* **31**, 745-759 (2020).
5. J. Li *et al.*, Clinical applications of liquid biopsy as prognostic and predictive biomarkers in hepatocellular carcinoma: circulating tumor cells and circulating tumor DNA. *J Exp Clin Cancer Res* **37**, 213 (2018).
6. H. T. Nguyen *et al.*, Ultra-Deep Sequencing of Plasma-Circulating DNA for the Detection of Tumor- Derived Mutations in Patients with Nonmetastatic Colorectal Cancer. *Cancer Invest* **40**, 354-365 (2022).
7. J. Pascual *et al.*, ESMO recommendations on the use of circulating tumour DNA assays for patients with cancer: a report from the ESMO Precision Medicine Working Group. *Ann Oncol* **33**, 750-768 (2022).
8. H. T. Nguyen *et al.*, Evaluation of a Liquid Biopsy Protocol using Ultra-Deep Massive Parallel Sequencing for Detecting and Quantifying Circulation Tumor DNA in Colorectal Cancer Patients. *Cancer Invest* **38**, 85-93 (2020).
9. Q. Gao *et al.*, Circulating cell-free DNA for cancer early detection. *The Innovation* **3**, 100259 (2022).
10. N. Constantin, A. A. Sina, D. Korbie, M. Trau, Opportunities for Early Cancer Detection: The Rise of ctDNA Methylation-Based Pan-Cancer Screening Technologies. *Epigenomes* **6**, (2022).

11. T. H. Phan *et al.*, Circulating DNA methylation profile improves the accuracy of serum biomarkers for the detection of nonmetastatic hepatocellular carcinoma. *Future Oncol* **18**, 4399-4413 (2022).
12. D. Schrag *et al.*, Blood-based tests for multicancer early detection (PATHFINDER): a prospective cohort study. *Lancet* **402**, 1251-1260 (2023).
13. B. D. Nicholson *et al.*, Multi-cancer early detection test in symptomatic patients referred for cancer investigation in England and Wales (SYMPLIFY): a large-scale, observational cohort study. *Lancet Oncol* **24**, 733-743 (2023).
14. H. T. Nguyen *et al.*, Multimodal analysis of ctDNA methylation and fragmentomic profiles enhances detection of nonmetastatic colorectal cancer. *Future Oncol* **18**, 3895-3912 (2022).
15. T. M. Q. Pham *et al.*, Multimodal analysis of genome-wide methylation, copy number aberrations, and end motif signatures enhances detection of early-stage breast cancer. *Front Oncol* **13**, 1127086 (2023).
16. T. H. H. Nguyen *et al.*, Clinical validation of a ctDNA-Based Assay for Multi-Cancer Detection: An Interim Report from a Vietnamese Longitudinal Prospective Cohort Study of 2795 Participants. *Cancer Investigation* **41**, 232-248 (2023).
17. V. T. C. Nguyen *et al.*, Multimodal analysis of methylomics and fragmentomics in plasma cell-free DNA for multi-cancer early detection and localization. *Elife* **12**, (2023).
18. V. A. Adalsteinsson *et al.*, Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nature Communications* **8**, 1324 (2017).
19. A. Jamshidi *et al.*, Evaluation of cell-free DNA approaches for multi-cancer early detection. *Cancer Cell* **40**, 1537-1549.e1512 (2022).
20. T. H. H. Nguyen *et al.*, Clinical validation of a ctDNA-Based Assay for Multi-Cancer Detection: An Interim Report from a Vietnamese Longitudinal Prospective Cohort Study of 2795 Participants. *Cancer Invest*, 1-17 (2023).
21. Y. Zhang *et al.*, Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. *Nat Commun* **12**, 11 (2021).
22. M. A. Kerachian, M. Azghandi, S. Mozaffari-Jovin, A. R. Thierry, Guidelines for pre-analytical conditions for assessing the methylation of circulating cell-free DNA. *Clin Epigenetics* **13**, 193 (2021).
23. M. Dameri *et al.*, Standard Operating Procedures (SOPs) for non-invasive multiple biomarkers detection in an academic setting: A critical review of the literature for the RENOVATE study protocol. *Crit Rev Oncol Hematol* **185**, 103963 (2023).
24. D. Santini *et al.*, Network approach in liquidomics landscape. *J Exp Clin Cancer Res* **42**, 193 (2023).
25. H. Sung *et al.*, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* **71**, 209-249 (2021).
26. X. Hao *et al.*, DNA methylation markers for diagnosis and prognosis of common cancers. *Proc Natl Acad Sci U S A* **114**, 7414-7419 (2017).

27. P. Liu, Pan-Cancer DNA Methylation Analysis and Tumor Origin Identification of Carcinoma of Unknown Primary Site Based on Multi-Omics. *Front Genet* **12**, 798748 (2021).
28. M. J. O'Brien *et al.*, The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. *Gastroenterology* **98**, 371-379 (1990).
29. J. F. Hall, Management of Malignant Adenomas. *Clin Colon Rectal Surg* **28**, 215-219 (2015).
30. L. Zhang, Q. Lu, C. Chang, Epigenetics in Health and Disease. *Adv Exp Med Biol* **1253**, 3-55 (2020).
31. D. S. C. Han, Y. M. D. Lo, The Nexus of cfDNA and Nuclease Biology. *Trends Genet* **37**, 758-770 (2021).
32. A. M. Lennon *et al.*, Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. *Science* **369**, (2020).
33. E. A. Klein *et al.*, Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol* **32**, 1167-1177 (2021).
34. Y. Tao *et al.*, Cell-free multi-omics analysis reveals potential biomarkers in gastrointestinal cancer patients' blood. *Cell Rep Med* **4**, 101281 (2023).
35. C. Keup, R. Kimmig, S. Kasimir-Bauer, Combinatorial Power of cfDNA, CTCs and EVs in Oncology. *Diagnostics (Basel)* **12**, (2022).

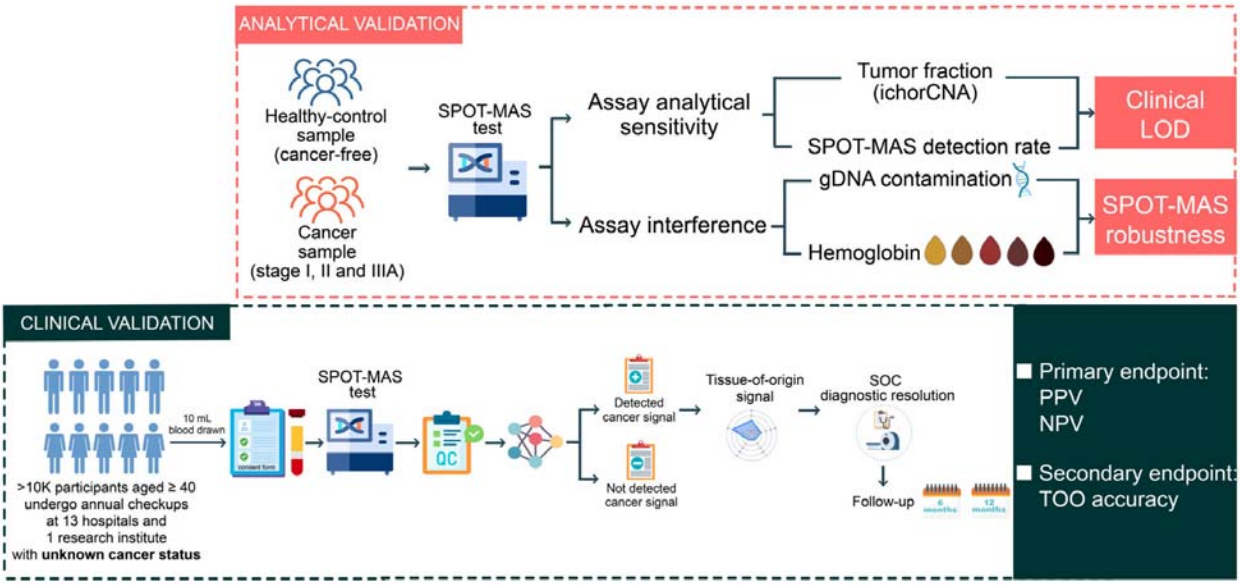

Figure 1. Overview of analytical and clinical validation.

For analytical validation, there are 3 investigations, including determination of limit of detection (LOD), potential interferences (genomic DNA and hemoglobin) and reproducibility. For clinical

validation, K-DETEK is a prospective study for previously unknown cancer status from 13 hospitals and one research institute in Vietnam. Plasma cfDNA was extracted from 10 ml blood of eligible participants. The cfDNA extraction was analyzed by SPOT-MAS assay. Using machine learning algorithms, SPOT-MAS provides two test results, including “detected cancer signal” or “not detected cancer signal” and tumor of origin. The participants with a detected cancer signal were consulted by physicians and confirmed in the diagnostic imaging tests according to the TOO probability values. All participants were followed up after 12 months to obtain information on possible cancer diagnosis.

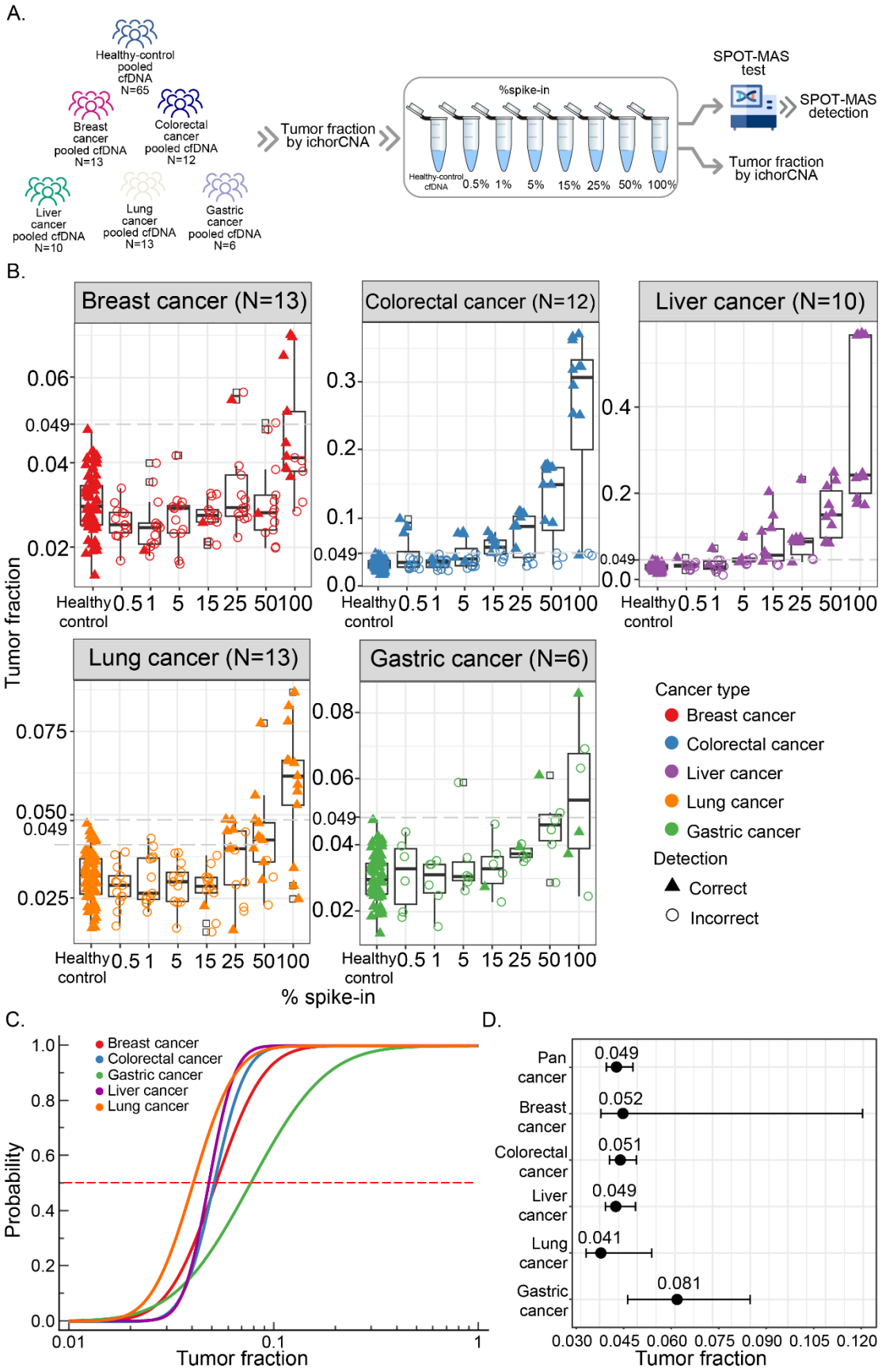
A.
Healthy-control pooled cfDNA N=65
Breast cancer pooled cfDNA N=13
Colorectal cancer pooled cfDNA N=12
Liver cancer pooled cfDNA N=10
Lung cancer pooled cfDNA N=13
Gastric cancer pooled cfDNA N=6
Tumor fraction by ichorCNA
%spike-in
Healthy-control cfDNA
0.5% 1% 5% 15% 25% 50% 100%
SPOT-MAS test
SPOT-MAS detection
Tumor fraction by ichorCNA
B.
Breast cancer (N=13)
Colorectal cancer (N=12)
Liver cancer (N=10)
Lung cancer (N=13)
Gastric cancer (N=6)
Tumor fraction
% spike-in
Healthy control
Cancer type
Breast cancer
Colorectal cancer
Liver cancer
Lung cancer
Gastric cancer
Detection
Correct
Incorrect
C.
Breast cancer
Colorectal cancer
Gastric cancer
Liver cancer
Lung cancer
Probability
Tumor fraction
D.
Pan cancer 0.049
Breast cancer 0.052
Colorectal cancer 0.051
Liver cancer 0.049
Lung cancer 0.041
Gastric cancer 0.081
Tumor fraction

Figure 2. Determination of limit of detection

A. Workflow describes a measure of limit of detection by SPOT-MAS test and tumor fraction-based ichorCNA.
B. Graphs show the correlation between the tumor fraction spiked-in percentage and tumor fraction estimation for long fragment in breast cancer (red), colorectal cancer (blue), gastric cancer (green), liver cancer (purple) and lung cancer (orange). Samples called as correct detected or incorrect detected are indicated by triangles or open circles, respectively.
C. Line chart shows the probability of five cancer types through tumor fraction at the probability of 50% cancer signal detected (red dotted line).
D. The LOD values of five cancer types using tumor fraction. Error bars denote 95% confidence intervals.

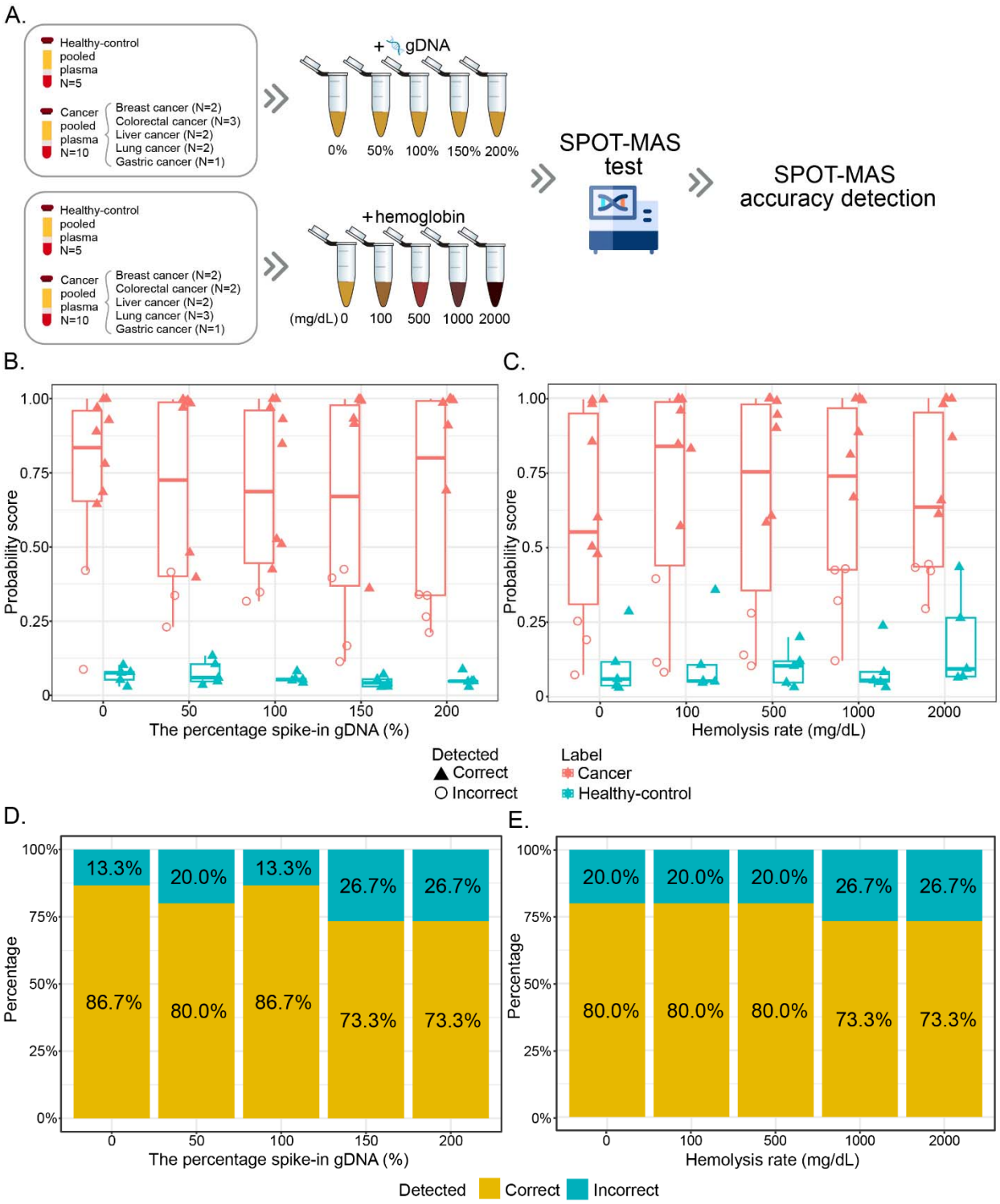

Figure 3. Effect of potential interferents on the performance of SPOT-MAS test

A. Workflow describes the preparation of potential interferent studies. Healthy-control plasma samples and cancer plasma samples were spiked with genomic DNA (0-200% of total cfDNA extracted from unspiked samples), hemoglobin (0-2000 mg/dL).

B-C. Boxes show probability scores in cancer and healthy-control samples spiked with 5 differents concentrations of genomic DNA (B) and hemoglobin (C). Samples called as correct detected or incorrect detected are indicated by triangles or open circles, respectively.

D-E. Charts show the percentage of detection in cancer and healthy-control samples spiked with 5 differents concentrations of genomic DNA (B) or hemoglobin (C).

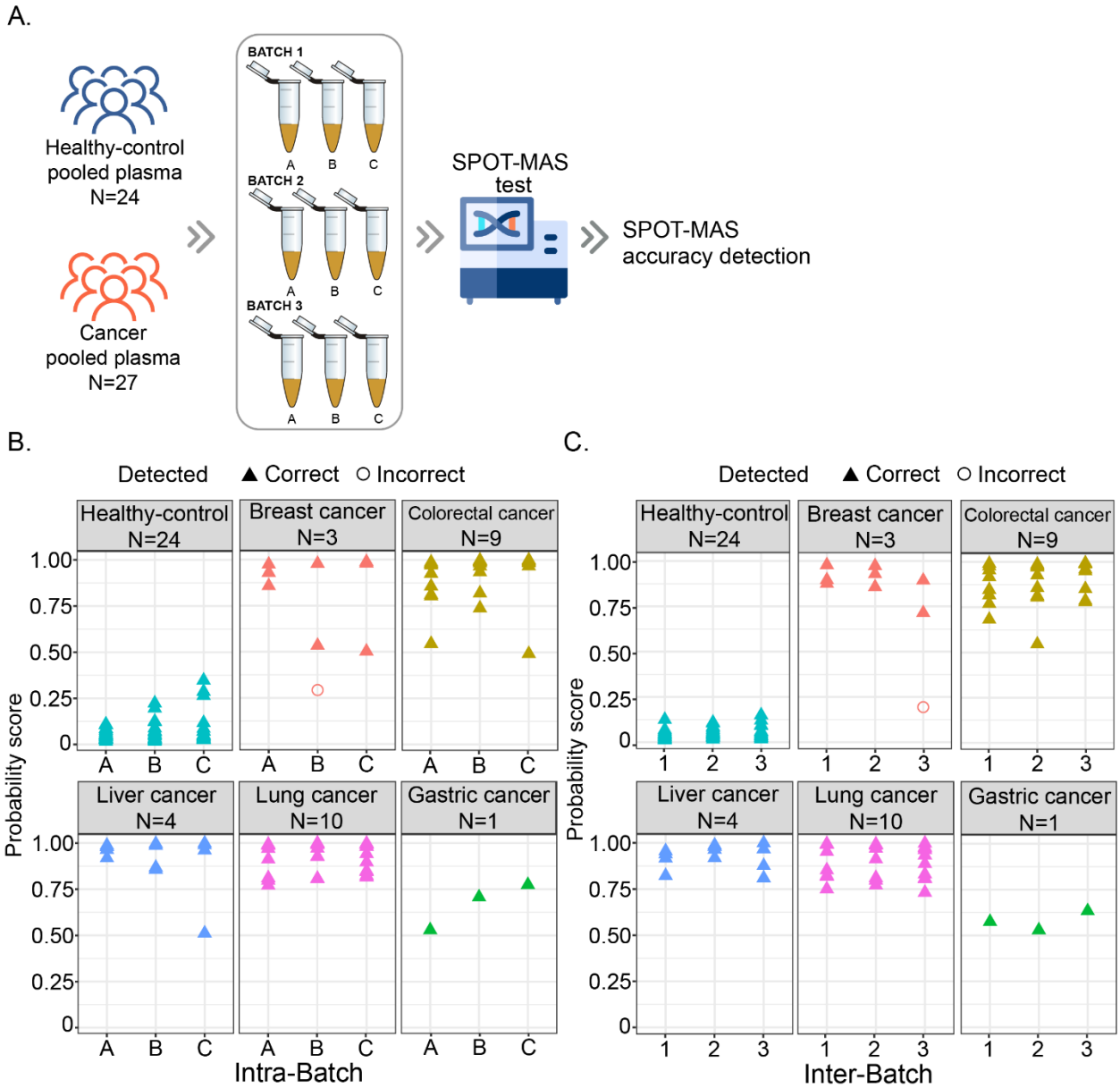

Figure 4. Reproducibility of SPOT-MAS test

A. Workflow shows experimental design of reproducibility study.

B-C. Boxes show probability scores performed intra-batch (B) and inter-batch (C) in breast cancer colorectal, gastric, liver, lung cancer and healthy control. Samples called as correct detected or incorrect detected are indicated by triangles or open circles, respectively.

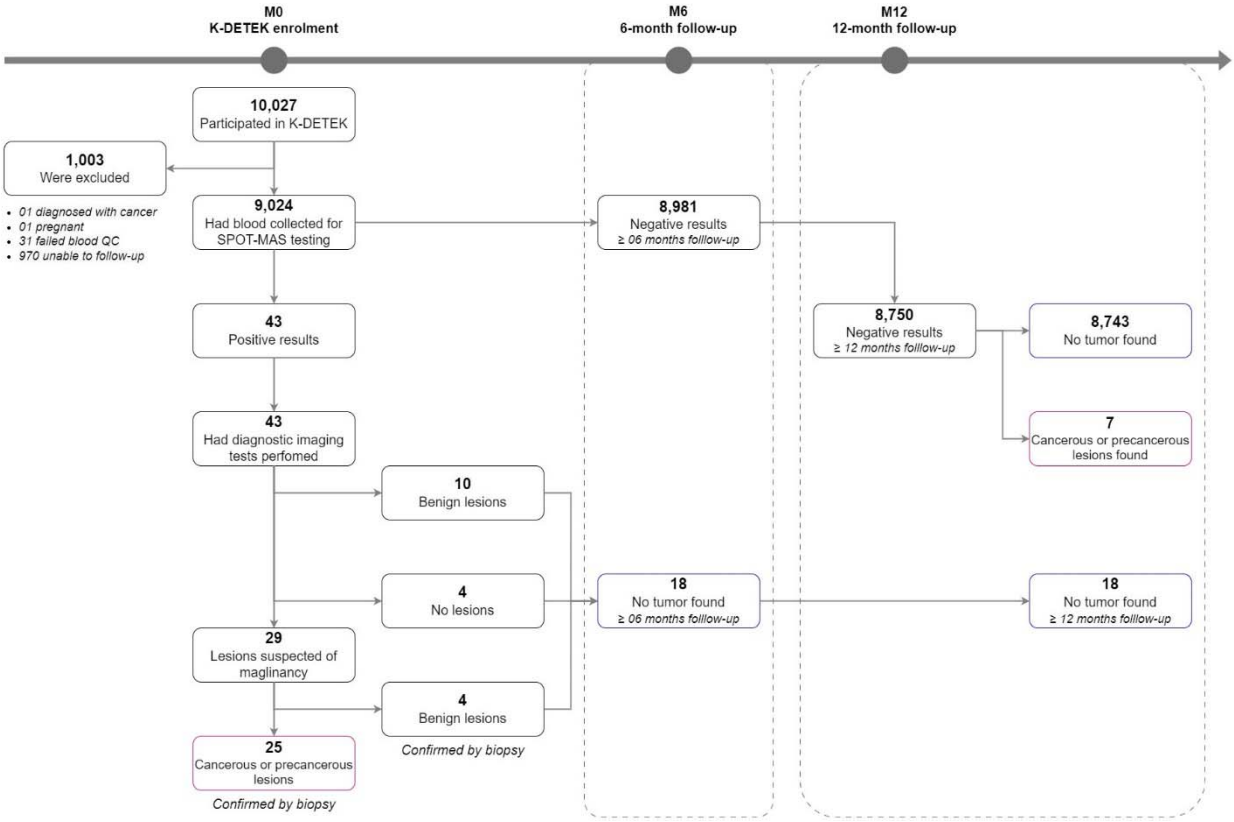

Figure 5. The flow chart of study subject selection and follow-up of K-DETEK.

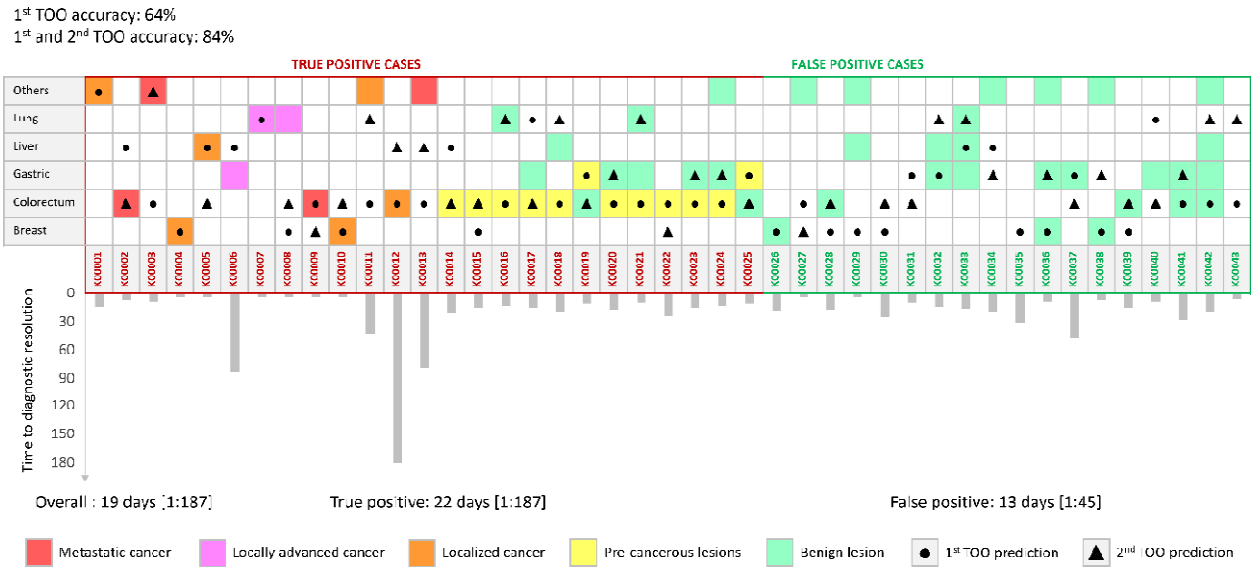

Figure 6. The determination of cancer specific origin prediction and cancer outcome.

Squares shows cancer specific origin of 43 cases diagnosed with a cancer signal detected using SPOT-MAS test. The bar charts show the observed time from receipt of positive ctDNA results to final diagnosis confirmation. Colors indicate the cancer outcomes, including metastatic cancer (red), locally advanced cancer (pink), localized cancer (orange), pre-cancerous lesions (yellow) and benign lesions (green).

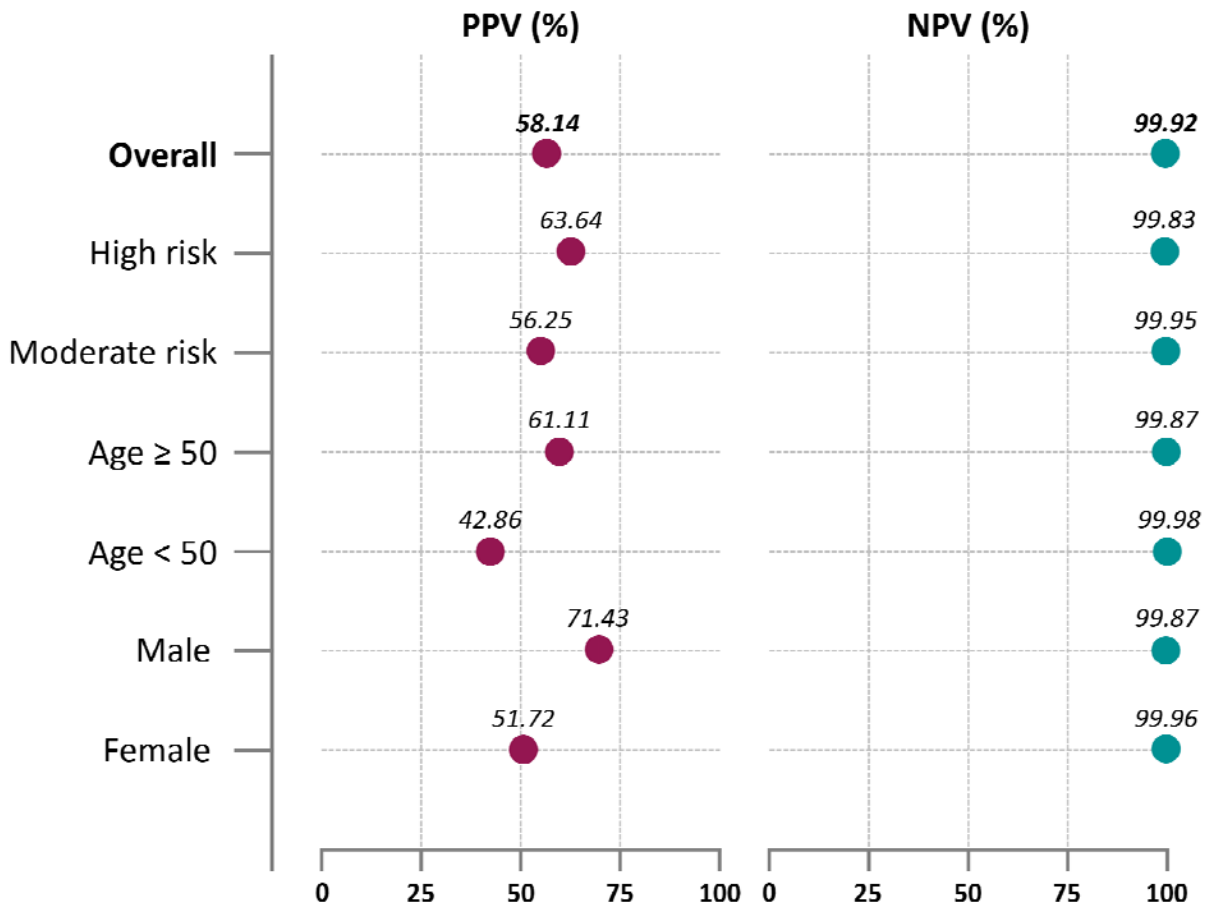

Figure 7. The effect of different clinical features on positive predictive value (PPV) and negative predictive value (NPV) of SPOT-MAS test.

Table 1. Clinical characteristics of 8,793 eligible participants

Risk factor		n=8,793	%
Age	<50	4,130	46.97
	>=50	4,663	53.03
	Median	50	
	Min	40	
	Max	94	
Gender	Female	4,800	54.59
	Male	3993	45.41
Liver infection	Yes	548	6.23
	No	8,245	93.77
Alcohol consumptions	Yes	501	5.70
	No	8,292	94.3
Smoking	Yes	465	5.29
	No	8,328	94.71
Hypertension	Yes	391	4.45
	No	8,402	95.55
Diabetes	Yes	146	1.66
	No	8,647	98.34
High risk		2,306	26.23
Moderate risk		6,487	73.77

Table 2. SPOT-MAS test performance

Test Performance		**N=8,793***	**%**
Detected		43	0.49
	True positive	25	0.28
	False positive	18	0.20
Not detected		8,750	99.51
	True negative	8,743	99.43
	False negative	7	0.08
PPV for cancer signal detection			58.14
NPV for cancer signal detection			99.92
Prediction accuracy of tumor origin			84.00

*Participants at 12-month follow-up