Analytical and clinical validation of a circulating tumor DNA-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], Nhu Nhat Tan Doan [1] , Hoang Tan Nguyen [1] , Phuoc Loc Doan [1] , Le Anh Khoa Huynh [1] , Tien An Nguyen [1] , Huu Tam Phuc Nguyen[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, methylation, fragment length.

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, the lack of analytical validation and clinical evidence for their utility in diverse populations have prevented their use in clinical practice. To address these challenges, we conducted a comprehensive analytical and clinical validation for an MCED test, SPOT-MAS (Screening for the Presence Of Tumor by DNA Methylation And Size). The analytical validation was to establish the limit of detection, reproducibility of test results and assess the impact of potential interferents on test performance. Specifically, SPOT-MAS could detect at least 50% of cancer samples at a specificity of 98% if the samples have tumor fraction ≥0.049 (95% CI: 0.043-0.059). The results were consistently reproduced for both intra- and inter-batch analysis. Moreover, our test remained robust at hemoglobin contamination of 500 mg/dl and genomic DNA contamination of up to 100%. To validate the performance of SPOT-MAS test in clinical settings, we launched a multi-center prospective trial, named K-DETEK, of 10,027 asymptomatic participants in Vietnam. Our assay achieved a positive predictive value of 58.14% (95%CI: 43.33-71.62) with 84.00% (95%CI: 65.35-93.60) accuracy in predicting tumor location and a negative predictive value of 99.92% (95%CI: 99.83-99.96). 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 an effort to combat this, conventional cancer screening methods, such as those recommended by the United States Preventive Services Task, have shown promise in increasing overall survival rate, improving treatment efficiency, and reducing long-term medical costs [3-5]. However, some conventional screening methods, particularly colonoscopy or cervical cytology tests, are invasive and have low accessibility [6,7]. Moreover, conventional screening tests only test for a single cancer type, leading to a high accumulative false positive rate and consequently overdiagnosis and overtreatment when performed sequentially [8,9]. One of the most promising advancements in cancer detection is the development of non-invasive multi-cancer early detection (MCED) tests [10]. MCED tests, using 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 [11]. LB assays detect specific cancer-related biomolecules including circulating tumour cells (CTC), cell free DNA (cfDNA), circulating tumor DNA (ctDNA), circulating free RNA (cfRNA) and exosomes [12-15]. 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 [12,13,16-18]. Recently, the landscape of MCED tests based on detecting methylation changes in cfDNA has evolved. The OverC test (Burning Rock) or Galleri test (Grail) can detect multiple cancer types simultaneously with high performance by interrogating methylation changes in cfDNA [19,20]. The Galleri test has been clinically validated in asymptomatic (PATHFINDER study) and symptomatic population

(SYMPLIFY study). The Galleri test was shown to exhibit a PPV of 43.1% for detecting cancer and the tumor of origin (TOO) prediction with an accuracy of 88% in an asymptomatic population [19]. Moreover, for individuals with malignancy-related symptoms, the SYMPLIFY trial in England and Wales demonstrated a particularly high PPV of 75.5% and NPV of 97.6% [20].

Despite promising results, 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 [11,21,22]. To improve the detection sensitivity of ctDNA, MCED screening methods tend to use high-depth sequencing, making it economically impractical for population-wide screening [11,23,24]. 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 origin of cancer signal [21,25]. By integrating cost-effective shallow sequencing and advanced machine learning, SPOT-MAS was trained and validated on a large cohort of 2,288 participants, including 738 nonmetastatic cancer patients and 1,550 healthy controls, achieving a sensitivity of 72.4% at 97.0% specificity and an accuracy of 70% for tumor-of-origin (TOO) prediction [26]. 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 [27]. In this study, we present the analytical validation of SPOT-MAS assay to determine the limit of detection and investigate the impact of technical factors on the robustness of our assay (**Figure 1**). Moreover, we recruited 10,027 asymptomatic participants in Vietnam to comprehensively evaluate the clinical performance of SPOT-MAS and demonstrate its clincal utility in the early detection of multiple types of cancer (**Figure 1**). The clinical applicability of SPOT-MAS is

measured by 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 and diagnostic-confirmed individuals for each type of cancer from our previous case-control study[26]. Healthy standard samples were collected from participants who were confirmed cancer-free at the time of enrollment and followed up for three years after enrollment to ensure that they did not develop cancer. For each cancer type including breast, colorectal, gastric, liver, and lung, standard samples were confirmed 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, 63 pooled plasma samples from healthy individuals and 50 pooled plasma samples from cancer patients, including breast cancer (n= 13), colorectal cancer (n= 12), liver cancer (n=

10), lung cancer (n= 9), and gastric cancer (n= 6), 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 [28]. 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. In this dtudy, 6 to 12 replicates of each TF level for each cancer type were investigated. The LOD of SPOT-MAS for each type of cancer was 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. [29].

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%) 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 were used. 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 annual health check-ups across 13 hospitals and one research institute in Vietnam from April 2022 to April 2023. 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 provided written informed consent prior to participating in this 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 3 months prior to recruitment time 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, ranging 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 previously [ref]. 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 motifs, were simultaneously analyzed by SPOT-MAS assay. The SPOT-MAS procedure was previously described [26]. Using machine learning algorithms, the model returned the

probability scores of ctDNA detection (SPOT-MAS score) 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 not detected" (negative) or "ctDNA signal detected" (positive) with up to two prediction results of TOO.

The participants with a detected ctDNA signal were consulted by physicians to undertake 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 are listed in **Table S2**. Participants with "other cancer" status were recommended for a whole-body CT scan for cancer screening. Participants with no cancer signal detected were informed about their lack of risk for the five cancer types covered by 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 S3**. 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 carriers. The remaining participants were classified in the moderate risk group.

All statistical analyses were performed by using R (4.3.2) with standard data analysis packages and the ggplot2 package for visualization. Confidence intervals were analyzed by Wilson method using R (4.3.2)[30]. Sigmoid curves with probit regression in LOD analysis were used software MedCalc. v22.001. Correlation analysis between tumor fractions estimated by ichorCNA and percentage of spike-in were measured by Pearson coefficient R.

RESULTS

The 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 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.[28] The ichorCNA 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, $R = 0.99$, $p = 1.7.10^{-11}$, **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 at different levels of the cancer cfDNA (0.5, 1, 5, 15, 25, 50 and 100%). As a result, for each cancer type we have from 6 to 13 pooled cancer cfDNA samples with known TF and we obtained seven spike-in cfDNA samples per pooled cancer cfDNA sample, giving us a total of 350 spike-in samples (**Table S4**). We observed significant positive correlations between TF levels and spike-in concentrations in most cancer types, with the highest correlation coefficient in liver cancer (R=0.84, $p< 2.2.10^{-6}$, **Figure 2B**), followed by colorectal cancer (R= 0.82, $p< 2.2.10^{-6}$, **Figure 2B),** lung cancer (R= 0.72, $p< 2.2.10^{-6}$, **Figure 2B**), gastric cancer (R= 0.57, $p=1.4.10^{-10}$, **Figure 2B**) and breast cancer (R=0.47, $p=6.5.10^{-10}$, **Figure 2B**). Thus, these results demonstrated the potential of the ichorCNA based approach in quantifying tumor fraction in plasma.

We next performed the SPOT-MAS test on 350 spike-in samples and 63 cfDNA samples from healthy individuals. We found that 63/63 (100%) of healthy samples were called correctly as non-cancer by the SPOT-MAS test, 129/350 (36.86%) were detected as cancer samples including 43/84 (51.19%) colorectal, 43/70 (61.43%) liver, 25/63 (39.68%) lung, 6/42 (14.29%) gastric and 12/91 (13.18%) breast cancer (**Table S5**). 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 0.043-0.059, **Figure 2D**). The lowest LOD was determined for lung cancer (0.038, 95% CI 0.027-0.098), followed by liver (0.049, 95% CI 0.043-0.059), colorectal (0.051, 95% CI 0.045-0.060), breast (0.052, 95% CI 0.041-0.18), 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 TF of greater than 0.049.

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

It has been well established that blood hemolysis causes genomic DNA (gDNA) and hemoglobin contamination which could interfere with the performance of liquid biopsy tests[31]. 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 gDNA or hemoglobin before running SPOT-MAS tests (**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, Table S6**). Likewise, the false negative rates were increased by 6.7% when the levels of hemoglobin contamination were greater than 500 mg/dL (**Figure 3E, Table S7**). Thus, our findings indicated that although the gDNA and hemoglobin contamination could affect the accuracy of SPOT-MAS performance, our test remained robust at the level of gDNA <100% and hemoglobin <500 mg/dL.

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 and 26/27 (96.30%) cancer samples were correctly detected by SPOT-MAS test across three different replicates (A, B and C); one breast cancer sample was incorrectly detected as healthy in replicate B (**Figure 4B, Table S8**). For the evaluation of inter-batch variations, we also observed similar levels of consistency, with 50/51 (98.04%) samples being correctly detected, and one breast sample in batch 3 being incorrectly detected (**Figure 4C, Table S8**). These findings showed a highly consistent performance of SPOT-MAS assay both within runs and between runs.

Evaluation of clinical performance of SPOT-MAS test

A total of 10,027 eligible participants were enrolled in the K-DETEK study from April 2022 to April 2023 from 13 hospitals and 1 research institute in Vietnam (**Figure 5**). Of those, 9,024 (90.00%) participants completed 6-month follow up and were included in the study analysis, while 1,003 participants (10.00%) were excluded from the study (**Table S3 and Figure 5**). 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 lost contact at 6-month follow-up period (n = 970) (**Figure 5**). A total of 8,793/9,024 (97.44%) participants completed the 12-month follow-up (**Figure 5**).

The clinical characteristics of the participants are summarized in **Table 1**, with a higher percentage of females than males (54.58% versus 45.42%, **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 26.24% (n = 2,307) of all participants. The remaining 73.76% (n = 6,486) individuals were considered moderate risk (**Table1**).

The majority of participants (8,750/8,793; 99.51%) showed “ctDNA signal not detected” results and 8,743 (99.92%) of them were confirmed to be cancer-free at 12 months after enrollment, indicating a NPV of 99.92% (95%CI: 99.83-99.96) (**Table 2 and Figure 5**). Of those, 7 (0.08%) cases were found to develop cancer during the 12-month follow-up (**Table S9**), with 2 cases developing metastatic lung cancer (n=2, patients K00044 and K00049), 2 cases with locally advanced colorectal cancer (n= 2, patients 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%, **Table 2 and Figure 5**) with "ctDNA signal detected" results, all of whom were referred to undertake on-site standard of care (SOC) imaging tests to confirm the presence of tumors according to our consultation protocol (**Table S2**). All cases agreed to undertake diagnostic tests for the cancer types corresponding to 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 (**Table 2 and Figure 5**). Of those cases, 25 were subsequently shown to have cancer or pre-cancerous lesions detected, suggesting a PPV of 58.14% (95%CI: 43.33-71.62) (**Table 2 and Figure 5**). 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% (95%CI: 65.35-93.60) for TOO (**Figure 6**). Specifically, 4 cases had metastatic colorectum, prostate, and cholangiocarcinoma (patients K00002, K00003, K00009, and K00013, **Figure 6 and Table S10**), 3 cases were diagnosed with locally advanced cancer at gastric and lung (patients K00006, K00007, and K00008, **Figure 6 and Table S10**), 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 and Table S10**). 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 and Table S10**). Four cases (patient K00001, K00003, K00011, and K00013, **Figure 6 and Table S10**) were found to have other tumor types that were not covered by SPOT-MAS 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 diagnostic results unable to confirm the

presence of malignant or precancerous tumors were regarded as false positive cases (**Table S10**). 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** and **Table S10**). 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 [32], 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% (95%CI: 39.33-71.83, **Figure 7**) in moderate risk participants to 63.64% (95%CI: 35.38-84.83, **Figure 7**) 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%, 95%CI 44.86-75.22 versus 42.86%, 95% CI: 15.82-74.95 **Figure 7**). For gender, the PPV was higher in male participants than female participants (71.43%, 95%CI: 45.35-88.28 versus 51.72%, 95%CI: 34.43-68.61, **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 PPV of 58.14% (95%CI: 43.33-71.62) and a NPV of 99.92% (95%CI: 99.83-99.96). Apart from detecting cancer signal, SPOT-MAS was able to identify the tissue localization of cancer signal with an accuracy of 84.00% (95%CI: 65.35-93.60), 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 successful treatment and improved survival outcomes for cancer patients [33]. 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 from 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 (median LOD 0.038), while showing the highest LOD value for detecting breast (median LOD 0.052) and 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 [26]. Our findings are also in line with previous studies showing low levels of ctDNA shedding by early-stage gastric and breast tumors [11,21,34].

Our previous findings showed the relationship between high hemolysis rate and its effect on isolated cfDNA and sequencing quality [31]. Therefore, we evaluated their potential interference 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 when the amounts of contaminants were greater than 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 [35-37], thereby may introducing confounding signals to our test. Therefore, the quality of blood samples should be meticulously controlled to ensure the robustness of the test.

The SPOT-MAS workflow was built based on genome-wide low-pass sequencing to maximize its cost effectiveness [26]. However, the use of low sequencing depth might introduce larger variation compared to high sequencing depth, which might lead to reduced reproducibility of the test [24,38]. We therefore designed the experiment to examine SPOT-MAS reproducibility. We showed that SPOT-MAS assay exhibited high reproducibility for the same samples tested either in the same sequencing runs (intra-batch) or in different sequencing runs (inter-batch) (**Figure 4**). This characteristic could be attributed to the design of our assay which simultaneously interrogated multiple features of cfDNA to achieve a final ctDNA probability score, thus alleviating the variability that might be associated with a particular feature [26].

The K-DETEK study was designed to determine the clinical performance of SPOT-MAS in a large asymptomatic cohort in Vietnam with the primary endpoint including the report of NPV and PPV, and the secondary endpoint is the report of TOO accuracy (**Figure 1**). Our data reported a NPV of 99.92% (95%CI: 99.83 - 99.96), corresponding to 8,743 out of 8,750 negative cases that remained cancer-free at 12 months after enrollment (**Table 2**). However, there were 7 participants who developed cancer within the 12-month follow-up period (ie. False negative cases). We found that 2 out of 7 cases (patient K00044 and K00050) developed metastatic cancer. Metastatic tumors are known to display methylation signatures different from primary early-stage tumors [39], and this may be the reason why these two cases were not detected by our algorithms, which were trained using samples with rigor selection criteria for non-metastatic

cancer (stage I-II and IIIA). Further analysis by ichorCNA showed that none of the 7 false negative cases displayed TF above the overall LOD of our test for detecting 5 cancer types (**Table S11 and Figure S2**). This finding indicated that all false negative cases in our study had low amounts of ctDNA in plasma beyond the capacity of SPOT-MAS test to detect.

Our findings from the K-DETEK study showed that SPOT-MAS test achieved a PPV of 58.14% (95%CI: 43.33-71.62, **Table 2**). For the 25 true positive cases, our analysis observed an accuracy of 84.00% (95%CI: 65.35-93.60, **Figure 6**) 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 (>= 50), male patients, or those belonging to high-risk group (**Figure 7**). 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. These polyps, including metaplasia and dysplasia, carry a risk of malignant cancer development ranging from 6% to 35% [40,41]. This underscores the critical role of SPOT-MAS in preventing the progression to malignant cancer, thereby improving clinical outcomes for cancer patients. Moreover, it only took 19 days after receiving SPOT-MAS results to achieve diagnostic resolution (**Figure 6**). Diagnostic resolution in such a short period of time could significantly alleviate patients' nervousness, and help expedite the necessary interventions for those diagnosed with precancerous conditions.

There were 18 cases who had ctDNA signal detected, but their diagnostic tests could not confirm the presence of a tumor (**Table 2**). It is possible that they might have early-stage tumors with small size that could not be detected by SOC imaging tests, and thus should have more regular check-ups. 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' [42,43].

In order to compare K-DETEK findings with other clinical studies, it is essential to consider the variations in population risk and their implications on test performance [19,20,44]. Our K-DETEK study revealed a positive rate of 0.28% (25 true positive in 8,793 cases), higher than the cancer incident rate in Vietnamese population (0.15%) [45], noting that the incident rate was estimated using all cancer types but with detection methods much different from our SPOT-MAS test. Our study focused on moderate- and high-risk participants with elevated chance of developing cancer, therefore 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 [46]. Moreover, our study observed a lower PPV compared to SYMPLIFY, which studied symptomatic patients with median age of 61.9 years [20]. Interestingly, 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 comparison to the GALLERI test used in the PATHFINDER study, which primarily focused on methylation makers [26,47].

Our study has several limitations. First, the determination of limit of detection by ichorCNA tool, which relies on CNA signals, might not provide an accurate reflection of the true LOD of SPOT-MAS test, particularly for the cancer types with low CNA signals. Second, a small proportion of participants (231/9024, 2.56%) did not complete the 12-month follow-up, which might have an impact on the estimation of NPV. Third, our assay has lower sensitivity for 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 a broader spectrum of cancer types, future studies are needed to explore the multi-omics approach that combine different makers, such as cfDNA, cfRNA and circulating tumor cells [48,49]. Fourth, the 18 false positive cases (0.2%) in this study could lead to unnecessary diagnosis procedure and indeed 14 experienced invasive endoscopy tests. However, SPOT-MAS has demonstrated promising results in maintaining a fixed PPV of 58.14% (95%CI: 43.33-71.62) 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. 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 that SPOT-MAS is a promising multi-cancer blood test for early cancer detection. Through comprehensive analytical validation, we have established the robustness of SPOT-MAS test, with a limit of detection of 0.049 (95% CI: 0.043-0.059) tumor fraction. Importantly, the data from K-DETEK study, the largest prospective trial in Asia, highlights the potential utility of SPOT-MAS in low- and middle-income countries. Together, our study offers compelling evidence for the feasibility of SPOT-MAS in early detection of asymptomatic cancer patients, thereby enhancing overall population health outcomes.

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:

The analytical data are available upon reasonable request by email to the corresponding authors (LST, MDP). Raw sequencing data are not publicly available due to ethical and regulatory restrictions.

Competing interests:

LHDN, THHN, NMP, BLT, THGN, DHV, THT, TDN, VTCN, YTL, THN, VUT, MPL, TMTT, MNN, TTVV, ANN, TTN, NNTD, HTN, PLD, LAKH, TAN, HTPN, CTTC, TVP, HST are employees of Gene Solutions. All other authors declare no competing interests.

Funding:

The study was funded by Gene Solutions.

Disclosure statement:

LST, HNN, HG, MDP, HHN and DSN hold equity in Gene Solutions. We confirm that this does not alter our adherence to the journal 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, , 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, 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, 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, Nhu Nhat Tan Doan , Hoang Tan Nguyen , Phuoc Loc Doan , Le Anh Khoa Huynh , Tien Anh 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: 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. We thank Gia Bao Nguyen, George School, PA, USA, for his help in proofreading this manuscript.

References

1. Pham T, Bui L, Kim G, et al: Cancers in Vietnam-Burden and Control Efforts: A Narrative Scoping Review. Cancer Control 26:1073274819863802, 2019
2. Hawkes N: Cancer survival data emphasise importance of early diagnosis. BMJ 364:l408, 2019
3. Davidson KW, Barry MJ, Mangione CM, et al: Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. Jama 325:1965-1977, 2021
4. Luo YH, Luo L, Wampfler JA, et al: 5-year overall survival in patients with lung cancer eligible or ineligible for screening according to US Preventive Services Task Force criteria: a prospective, observational cohort study. Lancet Oncol 20:1098-1108, 2019
5. Brill JV: Screening for cancer: the economic, medical, and psychosocial issues. Am J Manag Care 26:S300-s306, 2020
6. Joseph DA, King JB, Miller JW, et al: Prevalence of colorectal cancer screening among adults--Behavioral Risk Factor Surveillance System, United States, 2010. MMWR Suppl 61:51-6, 2012
7. Langsjoen J, Goodell C, Castro E, et al: Improving compliance with cervical cancer screening guidelines. Proc (Bayl Univ Med Cent) 28:450-3, 2015
8. Sasieni P, Smittenaar R, Hubbell E, et al: Modelled mortality benefits of multi-cancer early detection screening in England. Br J Cancer 129:72-80, 2023
9. Hackshaw A, Cohen SS, Reichert H, et al: Estimating the population health impact of a multi-cancer early detection genomic blood test to complement existing screening in the US and UK. Br J Cancer 125:1432-1442, 2021
10. Kakushadze Z, Raghubanshi R, Yu W: Estimating Cost Savings from Early Cancer Diagnosis, Data, 2017
11. Liu MC, Oxnard GR, Klein EA, et al: Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol 31:745-759, 2020
12. Li J, Han X, Yu X, 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
13. Nguyen HT, Luong BA, Tran DH, 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

14. Pascual J, Attard G, Bidard FC, 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

15. Nguyen HT, Tran DH, Ngo QD, 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

16. Gao Q, Zeng Q, Wang Z, et al: Circulating cell-free DNA for cancer early detection. The Innovation 3:100259, 2022

17. Constantin N, Sina AA, Korbie D, et al: Opportunities for Early Cancer Detection: The Rise of ctDNA Methylation-Based Pan-Cancer Screening Technologies. Epigenomes 6, 2022

18. Phan TH, Chi Nguyen VT, Thi Pham TT, 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

19. Schrag D, Beer TM, McDonnell CH, 3rd, et al: Blood-based tests for multicancer early detection (PATHFINDER): a prospective cohort study. Lancet 402:1251-1260, 2023

20. Nicholson BD, Oke J, Virdee PS, 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

21. Pham TMQ, Phan TH, Jasmine TX, 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

22. Cohen JD, Li L, Wang Y, et al: Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science 359:926-930, 2018

23. Cristiano S, Leal A, Phallen J, et al: Genome-wide cell-free DNA fragmentation in patients with cancer. Nature 570:385-389, 2019

24. Stackpole ML, Zeng W, Li S, et al: Cost-effective methylome sequencing of cell-free DNA for accurately detecting and locating cancer. Nat Commun 13:5566, 2022

25. Nguyen HT, Khoa Huynh LA, Nguyen TV, et al: Multimodal analysis of ctDNA methylation and fragmentomic profiles enhances detection of nonmetastatic colorectal cancer. Future Oncol 18:3895-3912, 2022

26. Nguyen VTC, Nguyen TH, Doan NNT, et al: Multimodal analysis of methylomics and fragmentomics in plasma cell-free DNA for multi-cancer early detection and localization. Elife 12, 2023

27. Nguyen THH, Lu YT, Le VH, 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

28. Adalsteinsson VA, Ha G, Freeman SS, et al: Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nature Communications 8:1324, 2017

29. Jamshidi A, Liu MC, Klein EA, et al: Evaluation of cell-free DNA approaches for multi-cancer early detection. Cancer Cell 40:1537-1549.e12, 2022

30. Newcombe RG: Improved confidence intervals for the difference between binomial proportions based on paired data. Stat Med 17:2635-50, 1998

31. Nguyen THH, Lu YT, Le VH, 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

32. Kim ES, Scharpf RB, Garcia-Closas M, et al: Potential utility of risk stratification for multicancer screening with liquid biopsy tests. NPJ Precis Oncol 7:39, 2023

33. Callister MEJ, Crosbie EJ, Crosbie PAJ, et al: Evaluating multi-cancer early detection tests: an argument for the outcome of recurrence-updated stage. Br J Cancer 129:1209-1211, 2023

34. Zhang Y, Yao Y, Xu Y, et al: Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. Nat Commun 12:11, 2021

35. Kerachian MA, Azghandi M, Mozaffari-Jovin S, et al: Guidelines for pre-analytical conditions for assessing the methylation of circulating cell-free DNA. Clin Epigenetics 13:193, 2021
36. Dameri M, Cirmena G, Ravera F, 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
37. Santini D, Botticelli A, Galvano A, et al: Network approach in liquidomics landscape. J Exp Clin Cancer Res 42:193, 2023
38. Luo X, Wang Y, Zou Q, et al: Recall DNA methylation levels at low coverage sites using a CNN model in WGBS. PLoS Comput Biol 19:e1011205, 2023
39. Hao X, Luo H, Krawczyk M, et al: DNA methylation markers for diagnosis and prognosis of common cancers. Proc Natl Acad Sci U S A 114:7414-7419, 2017
40. O'Brien MJ, Winawer SJ, Zauber AG, et al: The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. Gastroenterology 98:371-9, 1990
41. Hall JF: Management of Malignant Adenomas. Clin Colon Rectal Surg 28:215-9, 2015
42. Zhang L, Lu Q, Chang C: Epigenetics in Health and Disease. Adv Exp Med Biol 1253:3-55, 2020
43. Han DSC, Lo YMD: The Nexus of cfDNA and Nuclease Biology. Trends Genet 37:758-770, 2021
44. Lennon AM, Buchanan AH, Kinde I, et al: Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. Science 369, 2020
45. Sung H, Ferlay J, Siegel RL, 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
46. Nadauld LD, McDonnell CH, 3rd, Beer TM, et al: The PATHFINDER Study: Assessment of the Implementation of an Investigational Multi-Cancer Early Detection Test into Clinical Practice. Cancers (Basel) 13, 2021
47. Klein EA, Richards D, Cohn A, 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
48. Tao Y, Xing S, Zuo S, et al: Cell-free multi-omics analysis reveals potential biomarkers in gastrointestinal cancer patients' blood. Cell Rep Med 4:101281, 2023
49. Keup C, Kimmig R, Kasimir-Bauer S: Combinatorial Power of cfDNA, CTCs and EVs in Oncology. Diagnostics (Basel) 12, 2022

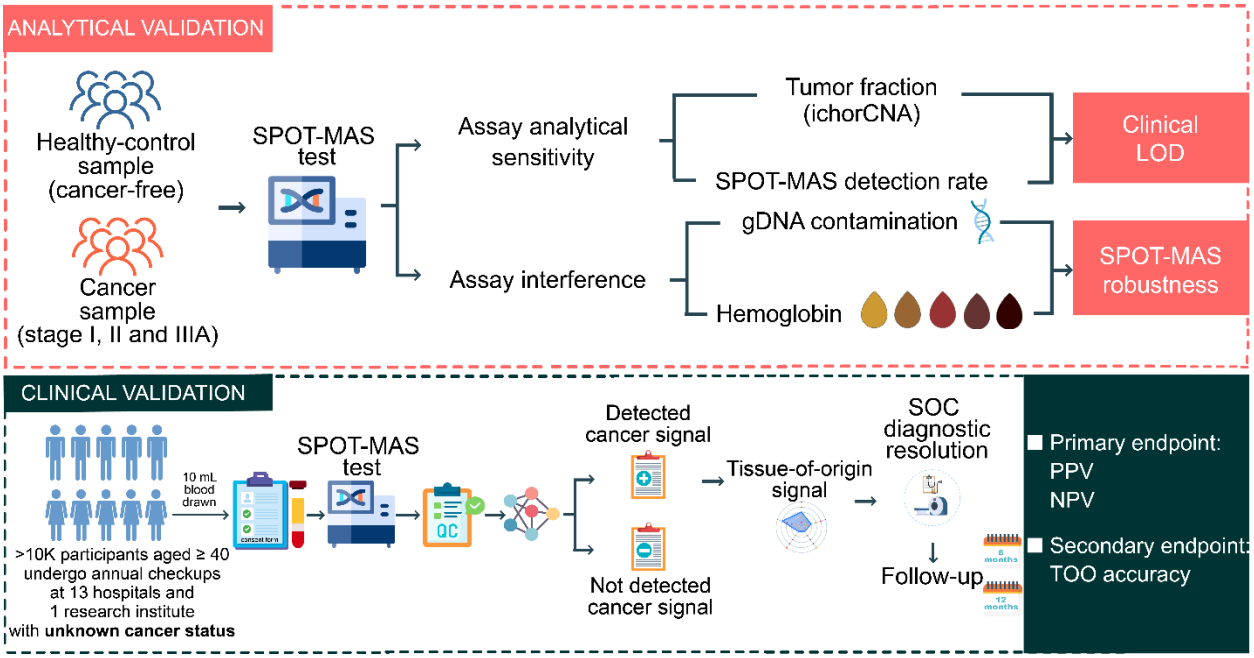

Figure 1. Overview of analytical and clinical validation.

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.

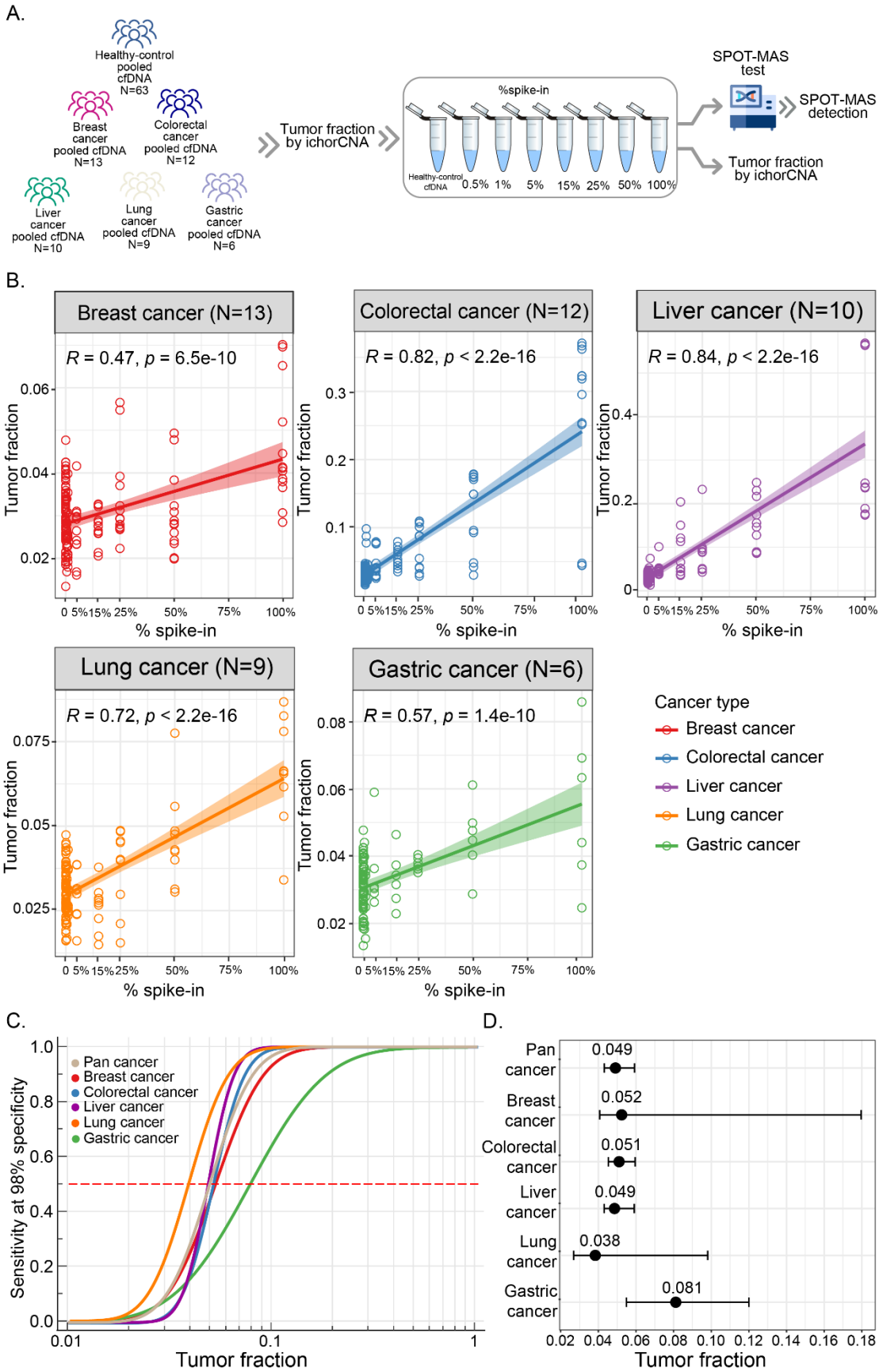
A.
Healthy-control pooled cfDNA N=63
Breast cancer pooled cfDNA N=13
Colorectal cancer pooled cfDNA N=12
Liver cancer pooled cfDNA N=10
Lung cancer pooled cfDNA N=9
Gastric cancer pooled cfDNA N=6
Tumor fraction by ichorCNA
%spike-in
Healthy-control cfDNA
0.5%
1%
5%
15%
25%
50%
100%
SPOT-MAS test
SPOT-MAS detection
Tumor fraction by ichorCNA
B.
Breast cancer (N=13)
R = 0.47, p = 6.5e-10
Colorectal cancer (N=12)
R = 0.82, p < 2.2e-16
Liver cancer (N=10)
R = 0.84, p < 2.2e-16
Lung cancer (N=9)
R = 0.72, p < 2.2e-16
Gastric cancer (N=6)
R = 0.57, p = 1.4e-10
Tumor fraction
% spike-in
Cancer type
Breast cancer
Colorectal cancer
Liver cancer
Lung cancer
Gastric cancer
C.
Pan cancer
Breast cancer
Colorectal cancer
Liver cancer
Lung cancer
Gastric cancer
Sensitivity at 98% specificity
Tumor fraction
D.
Pan cancer 0.049
Breast cancer 0.052
Colorectal cancer 0.051
Liver cancer 0.049
Lung cancer 0.038
Gastric cancer 0.081
Tumor fraction

Figure 2. Determination of limit of detection

A. Workflow describes a measure of tumor fraction-based on ichorCNA and limit of detection (LOD) by SPOT-MAS test
B. Graphs show the correlation between the spike-in percentage and tumor fraction estimation in breast cancer (red), colorectal cancer (blue), gastric cancer (green), liver cancer (purple) and lung cancer (orange). Correlation was assessed by Pearson coefficient R
C. Sigmoid curve shows limit of detection of 5 cancer types. LOD values are defined as as the TF levels at which the probability of detecting a cancer signal is at least 50% while maintaining at least 98% specificity (red dotted line).
D. The LOD values based on tumor fraction of five cancer types. Error bars denote 95% confidence intervals.

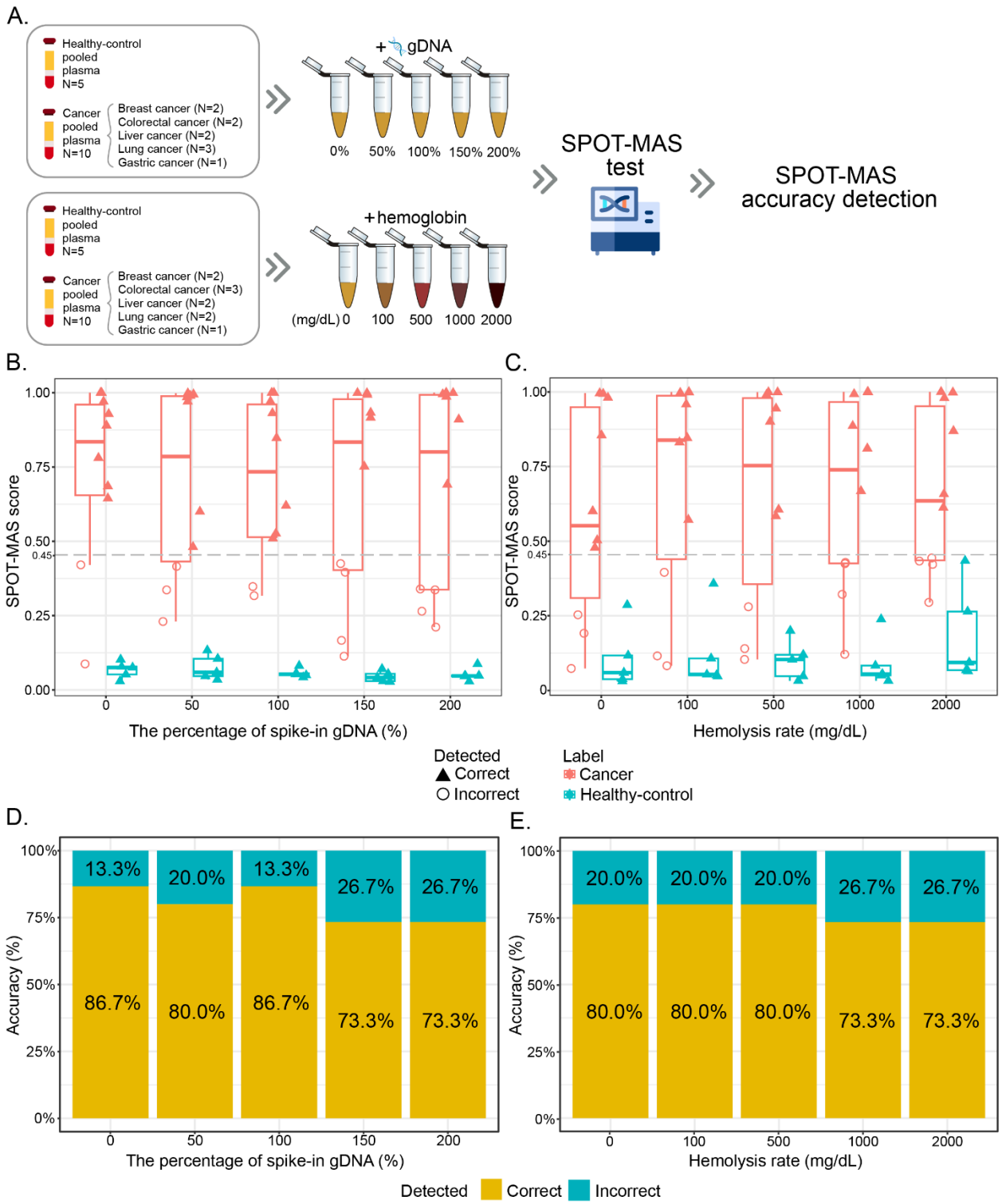

Figure 3. Effect of potential interferents on the performance of SPOT-MAS test

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). All samples were subjected to SPOT-MAS test to assess the accuracy of detection.

B-C. Box plots show SPOT-MAS scores in cancer (red) and healthy-control (cyan) samples spiked with 5 differents concentrations of genomic DNA (gDNA) (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 accuracy (%) in detecting cancer and healthy-control samples spiked with 5 differents concentrations of genomic DNA (B) or hemoglobin (C).

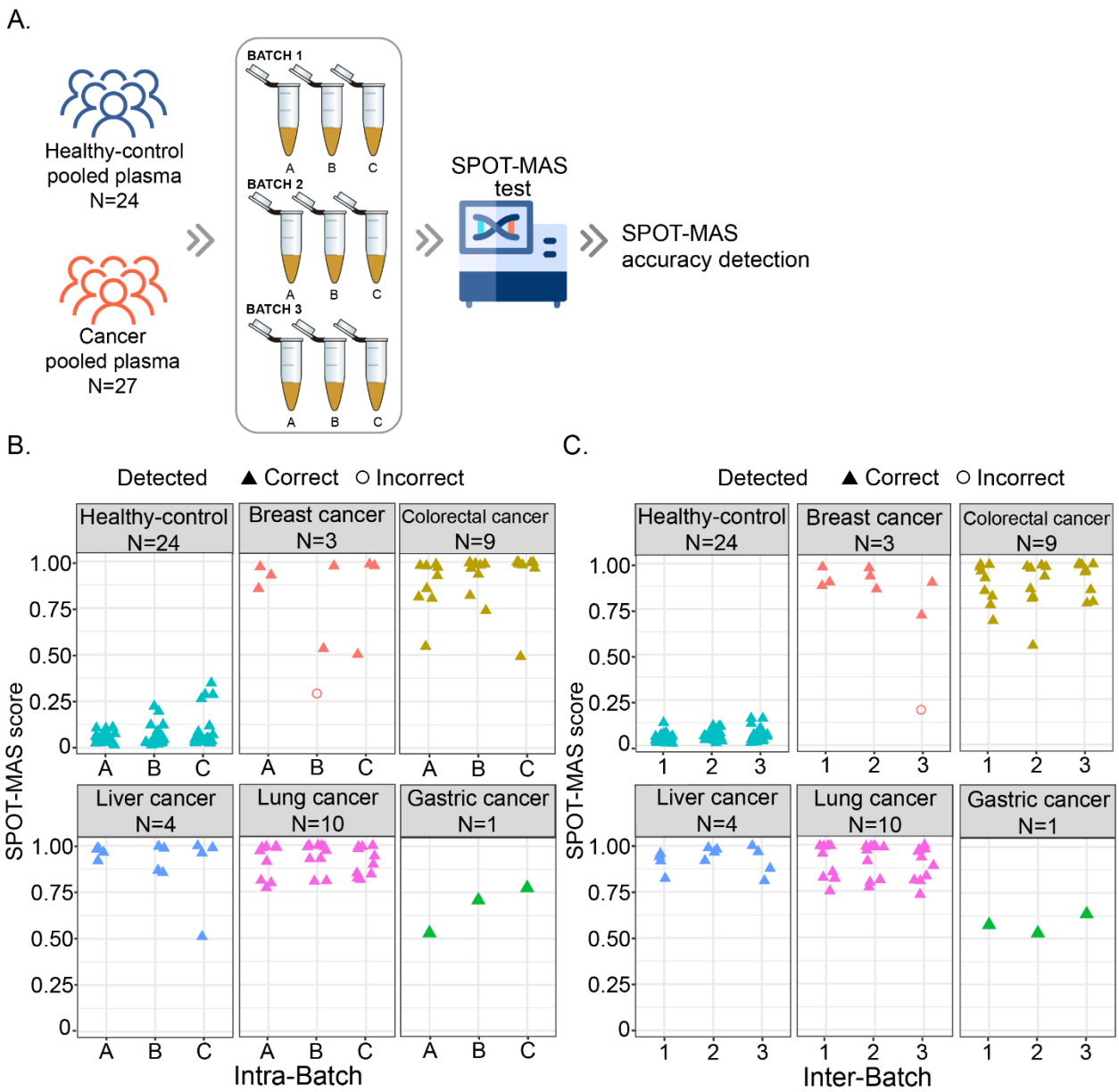

Figure 4. Reproducibility of SPOT-MAS test

A. Workflow shows experimental design of reproducibility study.

B-C. SPOT-MAS scores from intra-batch (B) and inter-batch (C) for breast (n=3), colorectal (n=9), gastric (n=1), liver (n=4), lung cancer (n=10), and healthy control (n=24). Samples called as correct detected or incorrect detected are indicated by triangles or open circles, respectively.

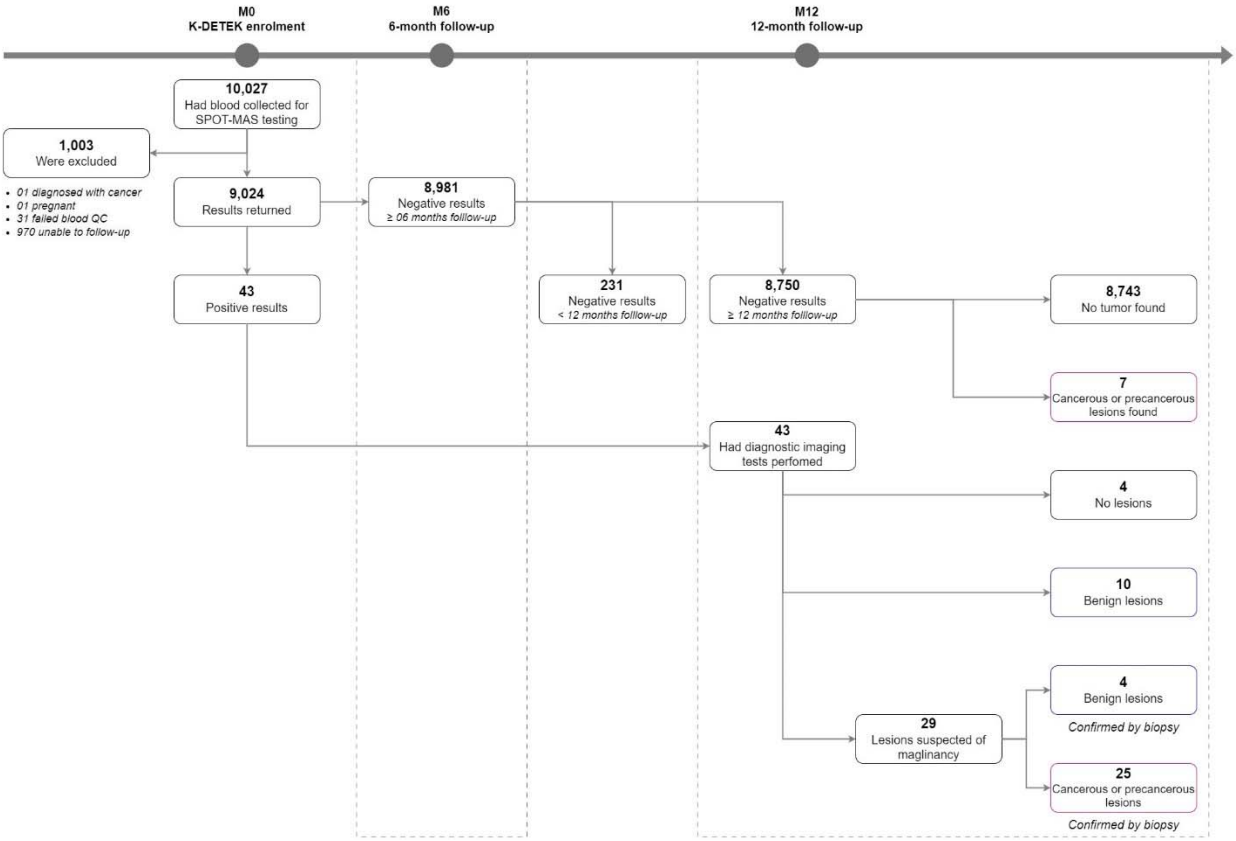

Figure 5. The flow chart of recruiting and following-up participants in the K-DETEK study

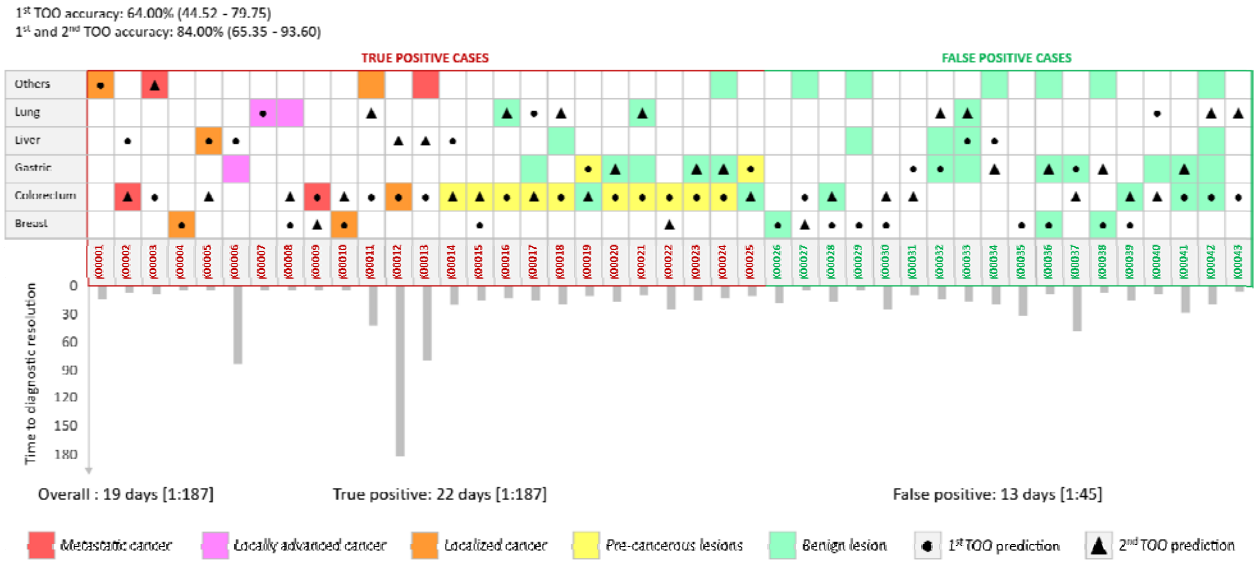

Figure 6. The analysis of diagnostic results of 43 participants with ctDNA signal detected.

Colored squares show lesion specific origin of 43 cases with cancer signal detected while circles and triangles denote the 1st and 2nd TOO prediction by SPOT-MAS. Colors indicate the cancer diagnostic outcomes, including metastatic cancer (red), locally advanced cancer (pink), localized cancer (orange), pre-cancerous lesions (yellow) and benign lesions (green). The intersections between colored squares and circles or triagles indicate the correct prediction of TOO by SPOT-MAS. The bar charts show the observed time from receipt of positive ctDNA results to final diagnosis confirmation.

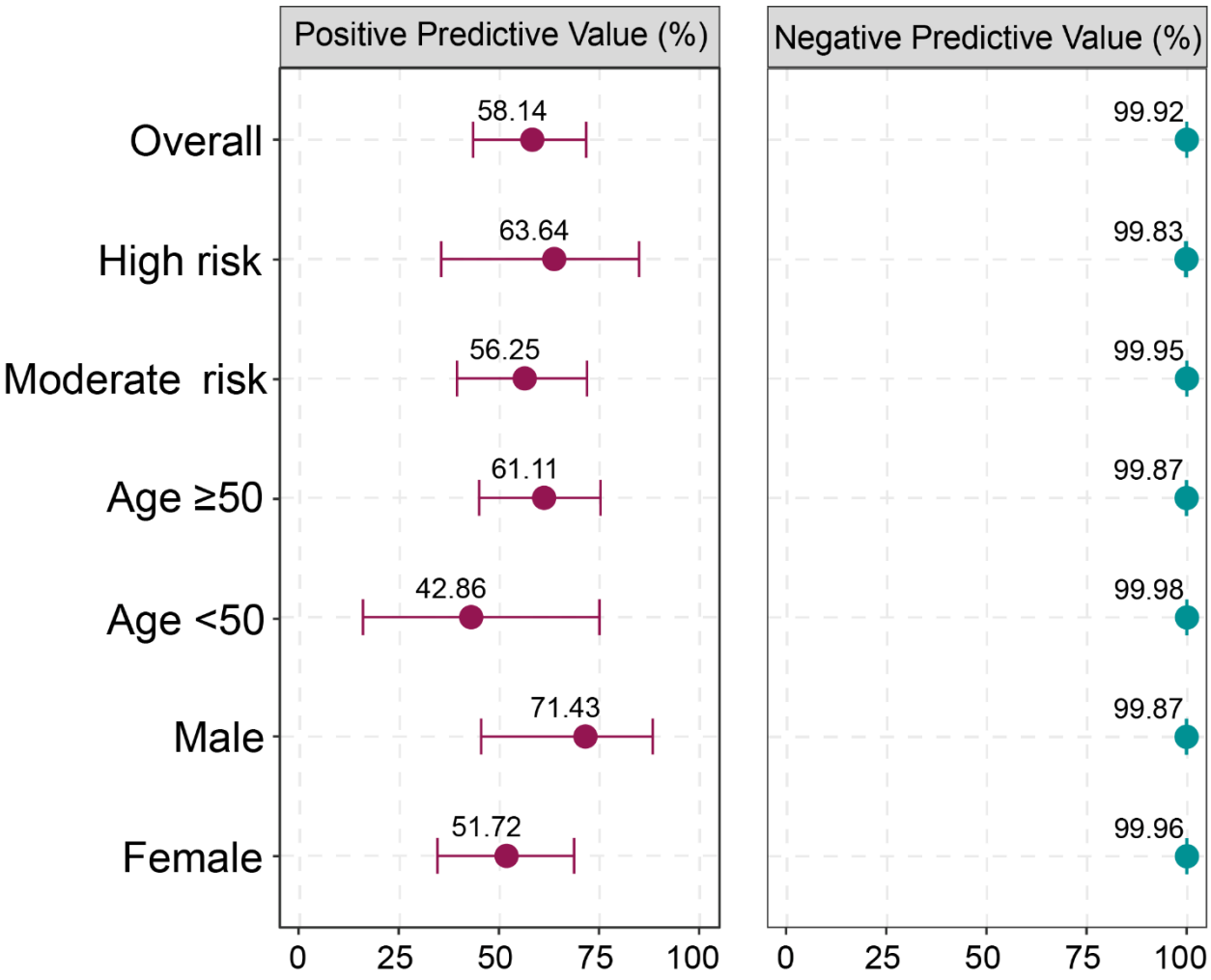

Figure 7. Positive predictive value (PPV) and negative predictive value (NPV) of SPOT-MAS test in different demographic groups.

Table 1. Clinical characteristics of 8,793 eligible participants

Risk factor		n=8,793	%
Age	<50	4,131	46.98
	>=50	4,662	53.02
	Median	50	
	Min	40	
	Max	94	
Gender	Female	4,799	54.58
	Male	3994	45.42
Liver infection (HBV/HCV)	Yes	548	6.23
	No	8,245	93.77
Alcohol consumptions	Yes	501	5.70
	No	8,292	94.30
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		2,307	26.24
Moderate risk		6,486	73.76

Table 2. SPOT-MAS test performance

Test Performance		**N=8,793***	**%**	**95% CI**
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	43.33 – 71.62
NPV for cancer signal detection			99.92	99.83 - 99.96
Prediction accuracy of tumor origin			84.00	65.35 – 93.60

*Participants at 12-month follow-up