Session A

62552825

Multiplatform analysis of early-stage cancer signatures in blood

Bingsi Li[1], Jiayue Xu[1] , Chenyang Wang[1], Fujun Qiu[1], Shuai Fang[1], Jing Su[1], Huiling Chu[1], Han Han-Zhang[1], Xinru Mao[1], Hao Liu[1], Heng Zhao[2#], Wei Zhang[3#], Xianling Liu[4#] , Zhihong Zhang[1#]

1. Burning Rock Biotech, Shanghai, China. 2. Shanghai Chest Hospital, Shanghai, China. 3. Changhai Hospital of Shanghai, Shanghai, China. 4. The 2nd Xiangya Hospital of Central South University, Changsha, China

#corresponding author. Email: zhihong.zhang@brbiotech.com

INTRODUCTION

Early detection of cancer is critical for effective treatment and reduced mortality. In patients with cancer, a portion of cell-free DNA (cfDNA) in the blood stream is tumor-derived (ctDNA), providing an opportunity to analyze the cancer genome noninvasively. Although profiling somatic mutations from ctDNA has achieved significant success for cancer diagnosis and surveillance, the sensitivity remains low for early-stage disease.

In a multi-center case-control study, we used a novel deep methylation sequencing technique called ELSA-seq to generate high-resolution maps of ctDNA from patients with cancers of lung, colorectum, or liver. To capture the highly diluted signals from early-stage tumors, a robust machine-learning classifier was built to identify cancer-associated signals and predict tissue of origin (TOO). Additionally, two somatic mutation profiling assays were evaluated in parallel in a pre-specified substudy.

MATERIALS AND METHODS

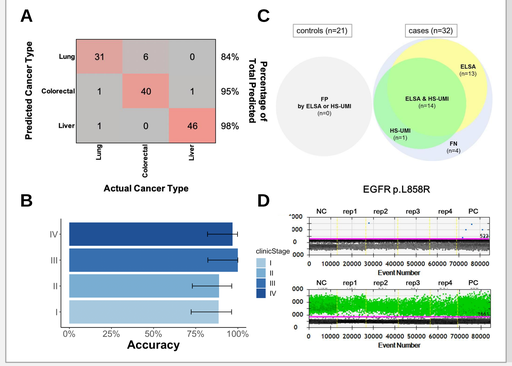
Study design:
The study was conducted among 490 patients with lung cancer (LC, N=178), colorectal cancer (CRC, N=187), liver cancer (LIHC, N=125), and 226 age-/sex- matched non-cancer controls. Surgery-resectable patients (stage I-III) consisted of 83%, 81%, and 86% for each cancer type respectively. Patients who were recognized to have anemia, acute infections, autoimmune diseases, or treated with neoadjuvant therapy were excluded from the study. The non-cancer controls were recruited with the criteria of showing no clinical symptoms or history of cancer at time of administration.

◊ **ELSA-seq:**
The target panel covers 80,672 CpG sites, spanning genomic regions of around 1.05Mb. In total, 8312 co-methylation blocks were defined and used as features/markers to build the classification model.

◊ **HS-UMI:**
The target panel spans 188kb of human genome, including 168 genes that are frequently mutated in lung cancer. Both cfDNA and paired WBC were sequenced with an average depth of 35,000X.

◊ **dd-PCR:**
QX200 (Bio-Rad) was used to absolutely quantify the copies of mutant and wild-type alleles in the sample. 30-50ng cfDNA input was required to reach 0.01-0.1% assay sensitivity.

FIGURE3



(A) Tissue of Origin (TOO) performance in the validation set. 92.6% (117/126) of the TOO predictions were correct.

(B) TOO of (A) categorized by stage. (stage I: n=35, stage II: n=36, stage III: n=23, stage IV: n=32)

(C) ELSA-seq detected 84% (27/32) of the cases and HS-UMI detected 47% (15/32) of those. Neither assay reported false positive (0/21).

(D) Detection of EGFR p.L858R mutation by ddPCR (found in primary tumor) at allele frequency (AF) of 0.02% for a patient who tested negative for HS-UMI but positive for ELSA-seq.

CONCLUSION

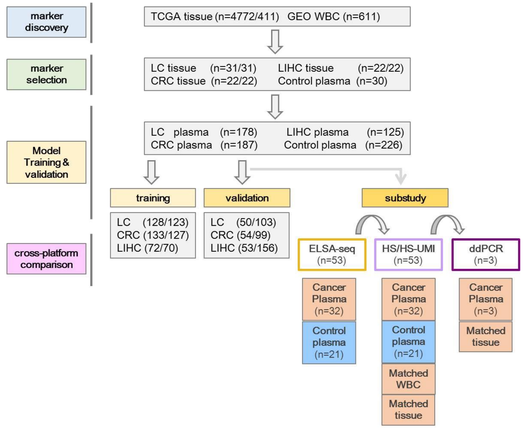
◊ At overall specificity of 95% (215/226), ELSA-seq demonstrated robust detection rates of 65% (86/133), 82% (125/153), 88% (101/115), and 94% (84/89) at stage I-IV respectively, highlighting the potential as a sensitive ctDNA profiling approach for early multi-cancer detection.

◊ ELSA-seq correctly predicted the TOO in 93% (117/126) of the cases in the validation set, offering an opportunity for decoding the ctDNA organ source.

◊ In a substudy, ELSA-seq identified nearly twice (27/32) as many patients as deep mutation sequencing (15/32) while reporting no false positives (0/21). It also required no biopsied tissues to detect ctDNA at AF as low as 0.02%.

◊ Large-scale prospective studies in high-risk populations and long-term follow-up will be needed for further evaluation .

FIGURE1



The study consists of four sequential steps: marker discovery, marker selection, model training and validation, and cross-platform evaluation. All samples were only used once at each step to avoid overfitting risks. To minimize batch effect, samples from case and control groups were mostly processed together.

A total number of 716 participants were included and randomly assigned into the training or validation groups according to preplanned ratios.

In a pre-specified substudy, 32 early-stage lung cancer patients and 21 non-cancer controls also had their cfDNA, white blood cells (WBC), and primary tumor samples analyzed by deep mutation sequencing (HS or HS-UMI), droplet digital PCR (ddPCR) singly or in combination. To unbiasedly evaluate the performance of different assays, ELSA-seq was set at 98% training specificity, similar to the other two methods reported in previous studies.

TABLE1

A

	None-cancer	Cancer	Colorectal Cancer	Liver Cancer	Lung Cancer
Total	226	490	187	125	178
Age, Mean+/-SD	56 +/- 6	59 +/- 8	61 +/- 8	55 +/- 8	61 +/- 6
Age, Min / Max	48 / 79	41 / 74	42 / 74	41 / 72	42 / 72
Sex, Female, n (%)	98 (43.4)	150 (30.6)	70 (37.4)	19 (15.2)	61 (34.3)
Clinical Stage, n (%)					
I	-	133 (27.1)	38 (20.3)	51 (40.8)	44 (24.7)
II	-	153 (31.2)	64 (34.2)	24 (19.2)	65 (36.5)
III	-	115 (23.5)	46 (24.6)	33 (26.4)	36 (20.2)
IV	-	89 (18.2)	39 (20.9)	17 (13.6)	33 (18.5)

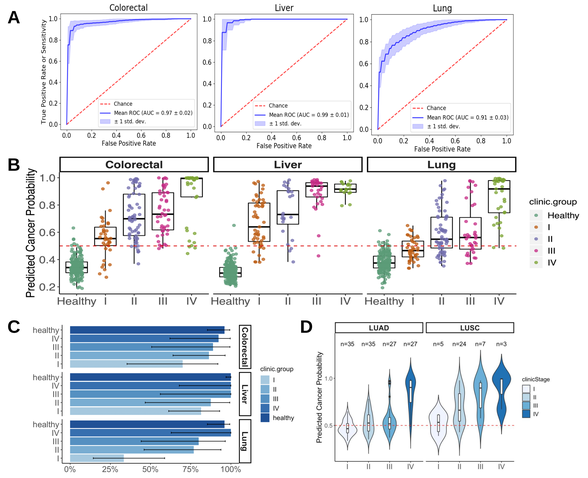
B

actual	predicted: stage	total	neg	pos	accuracy	95CI
LC	I	44	29	15	34.1%	20.9-50.0%
	II	65	18	47	72.3%	59.6-82.3%
	III	36	11	25	69.4%	51.7-83.1%
	IV	33	2	31	93.9%	78.4-98.9%
Non-cancer	-	226	217	9	96.0%	92.3-98.0%
CRC	I	38	9	29	76.3%	59.4-88.0%
	II	64	7	57	89.1%	78.2-95.1%
	III	46	2	44	95.7%	84.0-99.2%
	IV	39	3	36	92.3%	78.0-98.0%
Non-cancer	-	226	219	7	96.9%	93.5-98.6%
LIHC	I	51	9	42	82.4%	68.6-98.1%
	II	24	3	21	87.5%	66.5-96.7%
	III	33	1	32	97.0%	82.5-99.8%
	IV	17	0	17	100.0%	77.0-100%
Non-cancer	-	226	225	1	99.6%	97.2-100%
Sensitivity	I-IV	490	94	396	80.8%	77.0-84.1%
Specificity	-	226	215	11	95.1%	91.2-97.4%

(A) Clinical characteristics of the entire study cohort.

(B) ELSA-seq classification results of the entire cohort.

FIGURE2



(A) ROC curve and AUC from the training datasets. (CRC: 0.97, LIHC: 0.99, LC:0.91)

(B) Predicted cancer probability of the entire cohort categorized by disease stage.

(C) Classification results from the validation datasets (95% confidence intervals are indicated)

(D) Predicted cancer probability categorized by histological types. LUAD: lung adenocarcinoma, LUSC: lung squamous cell carcinoma