Molecular Cancer

CORRESPONDENCE

Open Access

Letter to the Editor: An ultra-sensitive assay using cell-free DNA fragmentomics for multi-cancer early detection

Hua Bao[1†], Zheng Wang[2,3,4†], Xiaoji Ma[5,6†], Wei Guo[7,8†], Xiangyu Zhang[2,3,4†], Wanxiangfu Tang[1], Xin Chen[1], Xinyu Wang[2,3,4], Yikuan Chen[5,6], Shaobo Mo[5,6], Naixin Liang[9], Qianli Ma[10], Shuyu Wu[1], Xiuxiu Xu[1], Shuang Chang[1], Yulin Wei[1], Xian Zhang[1], Hairong Bao[1], Rui Liu[1], Shanshan Yang[1], Ya Jiang[1], Xue Wu[1], Yaqi Li[5,6], Long Zhang[5,6,11], Fengwei Tan[7,8], Qi Xue[7,8], Fangqi Liu[5,6], Sanjun Cai[5,6,11*], Shugeng Gao[7,8*], Junjie Peng[5,6*], Jian Zhou[2,3,4,12,13*] and Yang Shao[1,14*]

Abstract

Early detection can benefit cancer patients with more effective treatments and better prognosis, but existing early screening tests are limited, especially for multi-cancer detection. This study investigated the most prevalent and lethal cancer types, including primary liver cancer (PLC), colorectal adenocarcinoma (CRC), and lung adenocarcinoma (LUAD). Leveraging the emerging cell-free DNA (cfDNA) fragmentomics, we developed a robust machine learning model for multi-cancer early detection. 1,214 participants, including 381 PLC, 298 CRC, 292 LUAD patients, and 243 healthy volunteers, were enrolled. The majority of patients ($N = 971$) were at early stages (stage 0, $N = 34$; stage I, $N = 799$). The participants were randomly divided into a training cohort and a test cohort in a 1:1 ratio while maintaining the ratio for the major histology subtypes. An ensemble stacked machine learning approach was developed using multiple plasma cfDNA fragmentomic features. The model was trained solely in the training cohort and then evaluated in the test cohort. Our model showed an Area Under the Curve (AUC) of 0.983 for differentiating cancer patients from healthy individuals. At 95.0% specificity, the sensitivity of detecting all cancer reached 95.5%, while 100%, 94.6%,

[†]Hua Bao, Zheng Wang, Xiaoji Ma, Wei Guo, and Xiangyu Zhang contributed equally to this work.

*Correspondence: caisanjuncsj@163.com; gaoshugeng@cicams.ac.cn; pengjj67@hotmail.com; zhou.jian@zs-hospital.sh.cn; yang.shao@geneseeq.com

[1] Geneseeq Research Institute, Nanjing Geneseeq Technology Inc., Nanjing 210000, Jiangsu, China
[2] Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai 200032, China
[5] Department of Colorectal Surgery, Fudan University Shanghai Cancer Center, Shanghai 200032, China
[7] Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China
[11] Department of Cancer Institute, Fudan University Shanghai Cancer Center, Fudan University, Shanghai 200032, China
Full list of author information is available at the end of the article

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

and 90.4% for PLC, CRC, and LUAD, individually. The cancer origin model demonstrated an overall 93.1% accuracy for predicting cancer origin in the test cohort (97.4%, 94.3%, and 85.6% for PLC, CRC, and LUAD, respectively). Our model sensitivity is consistently high for early-stage and small-size tumors. Furthermore, its detection and origin classification power remained superior when reducing sequencing depth to 1× (cancer detection: ≥ 91.5% sensitivity at 95.0% specificity; cancer origin: ≥ 91.6% accuracy). In conclusion, we have incorporated plasma cfDNA fragmentomics into the ensemble stacked model and established an ultrasensitive assay for multi-cancer early detection, shedding light on developing cancer early screening in clinical practice.

Keywords: Multi-cancer early detection, Cell-free DNA, Fragmentomics, Machine learning

Main text

The global cancer burden is increasing rapidly, and nearly 19.3 million new cases and 10.0 million cancer deaths were estimated in 2020 [1]. Over 60% of newly diagnosed cases and 70% of cancer mortality can be attributed to 10 common cancer types [1]. Among them, liver cancer, colorectal cancer, and lung cancer rank the top three causes and account for over one-third of cancer deaths [1]. Although cancer identified early is more likely to have a favorable prognosis [2], only limited early screening programs have been made available for specific cancer types [3]. Furthermore, detection limits, radiation exposure, fear of pain, monetary cost, etc., of existing screening programs are also obstacles in their implementation [4–6]. Therefore, exploring accurate and affordable biomarkers is needed for promoting early detection.

As a new class of biomarkers for cancer detection, cell-free DNA (cfDNA) in circulation is released from apoptosis and necrosis, and contains molecular signatures of its origin [7, 8]. For instance, tumor somatic mutations can serve as a classifier to distinguish circulating tumor DNA (ctDNA) shed from tumor cells and nontumorous cfDNA [9]. Epigenetic modifications such as DNA methylation and fragmentomic signatures such as fragmentation patterns and end motifs have also been utilized for identifying cancer [10–14]. However, assays based on single cfDNA features often yield inadequate detection ability, especially for stage I cases of prevalent cancer types [12, 14–16]. As identification at stage I often provides a better chance for the cure than later stages, developing more robust methods is critical to promote cancer early detection.

More recently, multi-dimensional predictive models that combine multiple fragmentomic and genomic features and even incorporate clinical information have improved their detection power for specific cancer types [12, 17]. Particularly, Ma et al. have leveraged the ensemble stacked strategy to integrate multiple fragmenomic features with machine learning algorithms and successfully built an ultrasensitive model for detecting stage 0/I colorectal adenocarcinoma [18]. Given the potential of the ensemble stacking strategy, we attempted to develop a multi-dimensional model using cfDNA fragmentomics from WGS data for multi-cancer detection and origin localization. Owing to their high prevalence and substantial impact, we built the model targeting liver, colorectal, and lung cancers in a cohort of the Chinese population. Our model demonstrated ultrasensitivity for cancer detection and accurately differentiated cancer origins, ideal for promoting cancer screening programs.

Results and discussion

Participant characteristics and disposition

We included 1,214 participants with previously untreated diseases: 381 primary liver cancer (PLC), 298 colorectal cancer (CRC), 292 lung adenocarcinoma (LUAD), and 243 healthy volunteers without cancer (Fig. 1A). This study was approved by the Ethnic Committees and in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments. Written informed consents were provided by all participants. Details about enrollment information are in Supplementary Materials and Methods. The participants were subject to WGS and fragmentomic feature extraction and randomly split into the training and test datasets

(See figure on next page.)

Fig. 1 Schematic diagram of the study design. **A** The training cohort ($N = 608$) included 191 primary liver cancer (PLC), 149 colorectal cancer (CRC), 146 lung adenocarcinoma (LUAD) patients, and 122 healthy controls, which were used to train the cancer detection and cancer origin models. The test cohort ($N = 606$), which included 190 PLC, 149 CRC, 146 LUAD, and 121 healthy controls, was used to evaluate model performances. **B** Plasma samples were collected from PLC, CRC, LUAD patients, and healthy volunteers. The cfDNA was extracted from the participant's plasma sample and subject to whole-genome sequencing (WGS). Five different feature types, including Fragment Size Coverage (FSC), Fragment Size Distribution (FSD), EnD Motif (EDM), BreakPoint Motif (BPM), and Copy Number Variation (CNV), were calculated. For each feature type, a base model was constructed based on the ensemble learning of five algorithms- GLM, GBM, Random Forest, Deep Learning, and XGBoost. The base model predictions were then ensembled into a large matrix, subsequently used to train the final ensemble stacked model

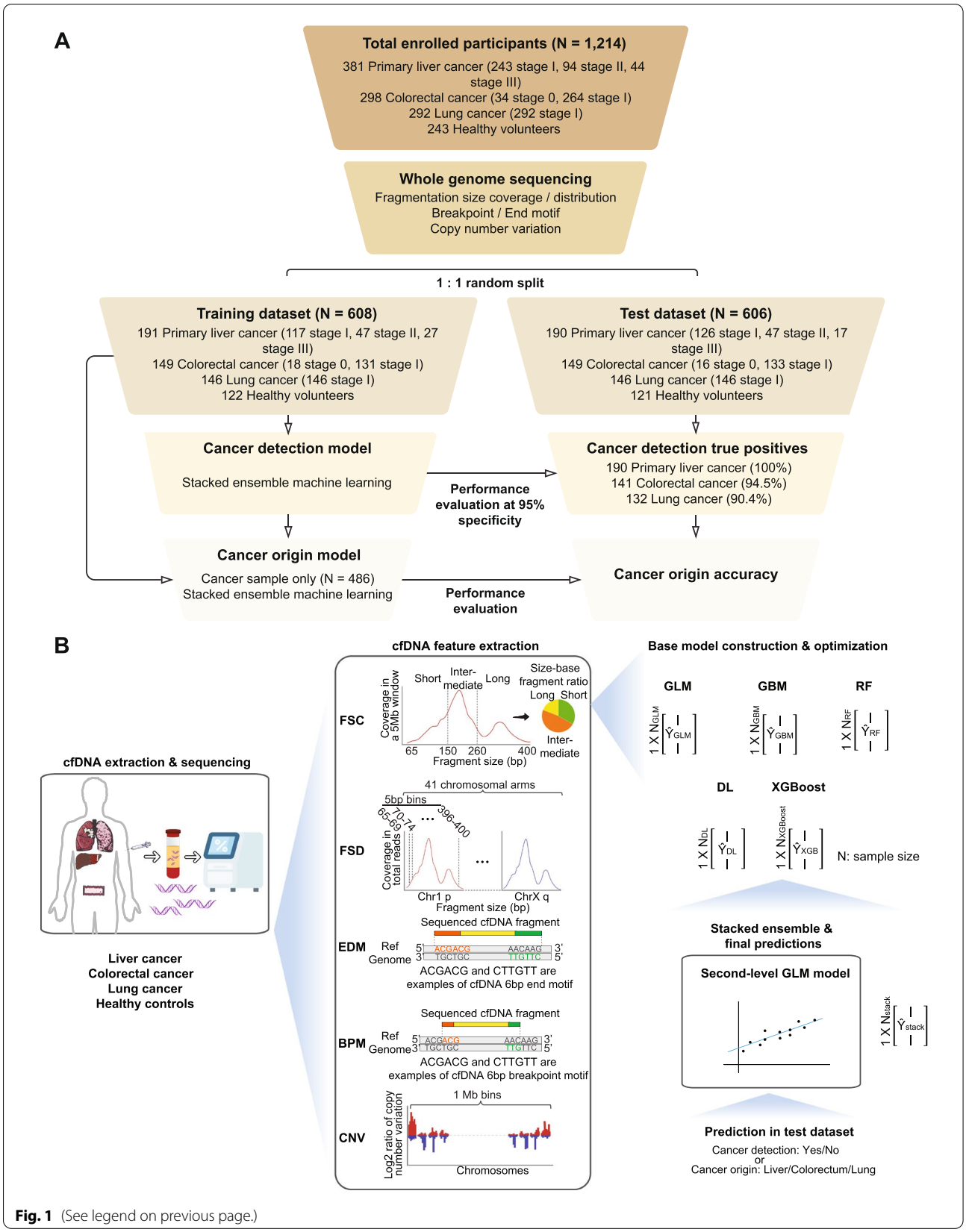


Fig. 1 (See legend on previous page.)

in a 1:1 ratio. We took the whole training dataset to build the first-level cancer detection model and then the cancer samples in the training dataset to train the second-level cancer origin model. The workflow of model construction is described in Fig. 1B and Supplementary Materials and Methods. Briefly, we extracted five distinct features covering cfDNA fragmentation size, motif sequence, and copy number variation from the WGS data, namely Fragment Size Coverage (FSC), Fragment Size Distribution (FSD), EnD Motif (EDM), BreakPoint Motif (BPM), and Copy Number Variation (CNV). The fragmentomic features implemented five machine learning algorithms, including Generalized Linear Model (GLM), Gradient Boosting Machine (GBM), Random Forest, Deep Learning, and XGBoost, and integrated to establish the ensemble stacked model. It is worth noting that the model was built solely in the training dataset, while the test dataset remained untouched until the model was finalized. We evaluated the cancer detection model in the test dataset and then took the true-positive cases to validate the cancer origin model. Healthy and cancer participants' demographics and characteristics (Table S1) are comparable between the training and test datasets. More importantly, the cancer samples are highlighted by the majority of early-stage diseases [PLC: stage IA/IB 117/191 (61.3%) in the training cohort and 126/190 (66.3%) in test cohort; CRC: stage 0/I 149/149 (100.0%); LUAD stage IA/IB 146/146 (100.0%)].

Differentiating cancer and non-cancer subjects by the cancer detection model

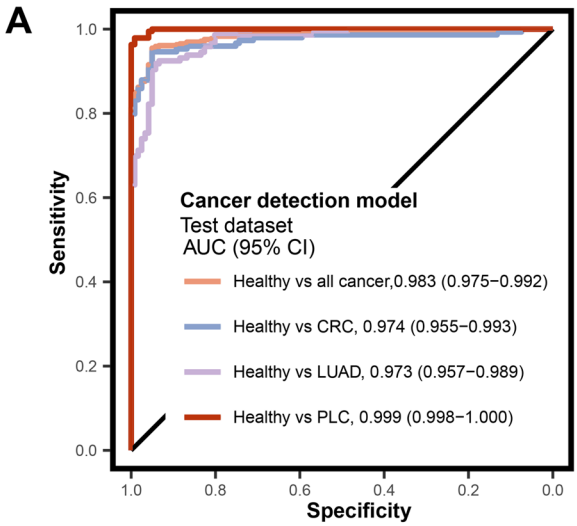
We reached a superior AUC value of 0.983 (95% CI: 0.975-0.992) for detecting all cancer subjects in the test dataset (Fig. 2A). The PLC group has the highest AUC (0.999, 95% CI: 0.975-0.992), followed by the CRC (0.974, 95% CI: 0.955-0.993) and LUAD (0.973, 95% CI: 0.957-0.989) groups. Healthy subjects have lower cancer scores than cancer subjects, and the three cancer types showed similar score distribution (Fig. 2B). The cancer score of 0.39 rendered a 95.0% specificity (95% CI: 89.5-98.2%). The corresponding sensitivities are 95.5% (95% CI: 93.2-97.1%) for all cancer subjects (Fig. 2C), and 100.0% (95% CI: 98.1-100.0%), 94.6% (95% CI: 89.7-97.7%), and 90.4% (95% CI: 84.4-94.7%) for PLC, CRC and LUAD, respectively (Table S2). We observed an upward trend from the early to later stages for the distribution of cancer scores in all-cancer, PLC, and CRC classes (Fig. S1). A propensity score matching analysis balanced the age and sex factors between cancer and non-cancer groups in the test dataset. The resultant subset consisting of 113 PLC, 73 CRC, 85 LUAD, and 85 age and gender-matched healthy controls remained high performance in distinguishing cancer patients from non-cancer controls (AUC: 0.988, 95% CI: 0.980-0.996, Fig. S2A). We also performed 10-fold cross-validation during training to evaluate model overfitting. The 10-fold cross-validation AUCs for all-cancer and individual cancer types were equally high compared to the independent test dataset (Fig. S2B), reassuring that overfitting was not a major concern.

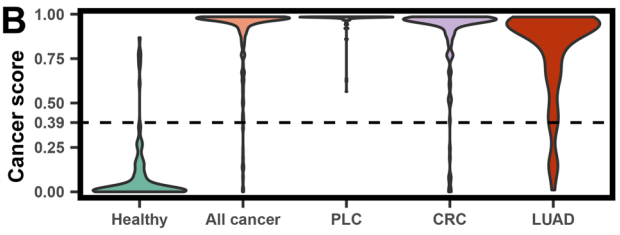
Our model exhibited ultrasensitivity in detecting cancers at various stages (Fig. 2D). The sensitivity is above 90% for stages 0 and I, and elevated to nearly 100% for stages II and III. Furthermore, we used patient demographics and clinical characteristics to categorize disease subgroups for evaluation (Table S3 and Figs. S3-S5). The model's detection sensitivity was consistently high even in the challenging categories, such as MIA and <1 cm tumors of LUAD. We assessed the model's robustness by gradually down-sampling the coverage to 1× (Fig. 2E Table S4). Despite a slight dip, the model remained stable with over 91.5% sensitivity for all-cancer. Even for the least detectable class of LUAD, the sensitivity at 1× is still above 87%.

Furthermore, the cancer detection model was assessed in a preliminary at-risk patient cohort and showed an overall specificity of 92.4% (Table S5, details in Supplementary Results).

(See figure on next page.)

Fig. 2 Performance and robustness valuation for the ensemble stacked model. **A** ROC curves evaluating the cancer detection model in distinguishing cancer patients from healthy volunteers in the test cohort, and further categorized into each cancer type class. **B** Violin plots illustrating cancer score distribution in the healthy, all cancer, primary liver cancer (PLC), colorectal cancer (CRC), and lung adenocarcinoma (LUAD) groups in the test cohort predicted by the cancer detection model. The 95% specificity cutoff for cancer score was 0.39, as shown by the dotted line. **C** Performance of the cancer detection model in identifying all cancer patients. **D** Dot plot of sensitivity in cancer detection by each cancer type and/or stage, at 95% specificity. The error bars represented the 95% confidence interval. **E** Robustness test for the cancer detection model using test cohort with downsampled coverage depth (4×-1×). The error bars were calculated based on five repeats for each coverage. **F** Confusion matrix of the selected test cohort by cancer detection model for the cancer origin model. **G** Violin illustrating cancer origin score distribution in the PLC, CRC, and LUAD groups in the selected test cohorts predicted by the cancer origin model. **H** Dot plot illustrating robustness test for the cancer origin model using the selected test cohort with downsampled coverage depth (4×-1×). **I** Heatmap illustrating the detailed results of each patient for the robustness test of the cancer origin model

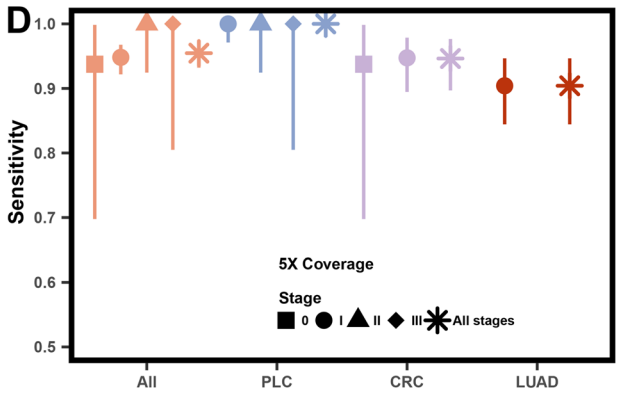


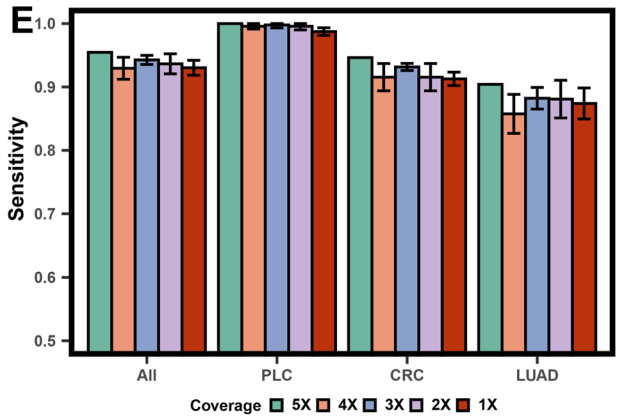


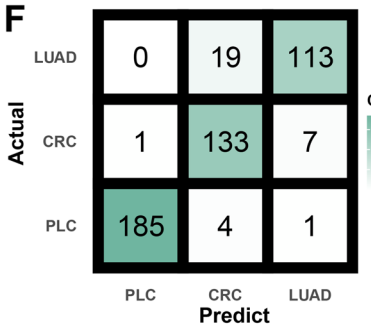
C

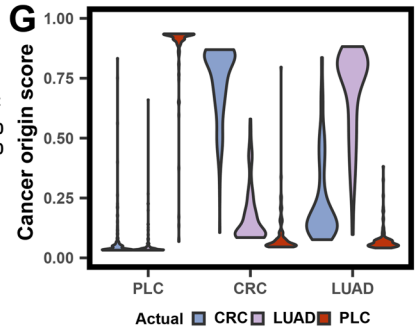
All cancer vs. healthy		Actual	
		All cancer	Healthy
Predict	All cancer	463	6
	Healthy	22	115
Sensitivity (95% CI)		95.5% (93.2-97.1%)	
Specificity (95% CI)		95.0% (89.5-98.2%)	

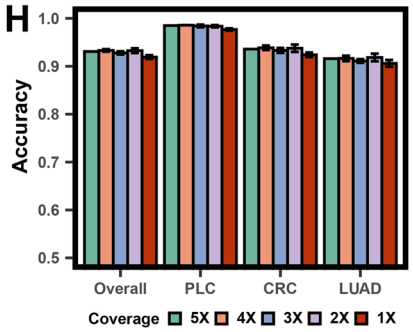
95% CI: 95% confidence interval











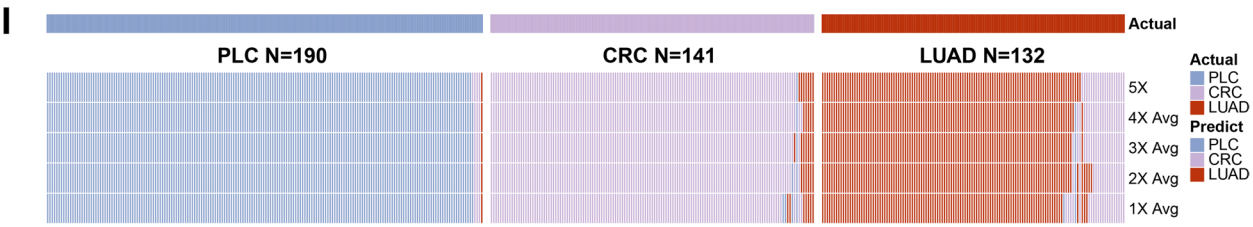


Fig. 2 (See legend on previous page.)

Locating cancer at its origin by the cancer origin model

All test dataset patients correctly identified as "Cancer" by the cancer detection model were subsequently analyzed in the cancer origin model. The model correctly identified the cancer origin for 431 patients (accuracy 0.931, 95% CI: 0.900-0.950) for the three cancer types (Fig. 2F and Table S6). The sensitivities for individual cancer types were 97.4% (95% CI: 94.0-99.1%), 94.3% (95% CI: 89.1-97.5%), and 85.6% (95% CI: 78.4-91.1%) for PLC, CRC, and LUAD, respectively. We plotted the cancer origin scores of each type for all patients (Fig. 2G). Generally, the top scores matched the true cancer types. Such consistency is the most compelling for the PLC patients, followed by the CRC patients, while the LUAD group has more erroneous CRC predictions (Fig. 2F and G). We further inspected the origin scores of the misinterpreted patients (Fig. S6). The score differences between the true origin and the misinterpreted class were minimal ($\leq$ 0.05) for potential improvement. The cancer origin model is robust with lower coverage WGS data (Fig. 2H and Table S7). The accuracies for PLC, CRC, and LUAD at 1× coverage are 97.7%, 92.4%, and 90.6%, respectively, whereas the predictions of each patient at different sequencing coverages were listed in Fig. 2I, H.

Our study has several limitations. First, we performed the proof-of-concept study using liver cancer, colorectal cancer, and lung cancer for their high prevalence. Targeting a broader population and more cancer types, including the less prevalent ones, would be necessary to develop the assay and eliminate cancer treatment inequity. Second, we are expanding our current cohort size to enable independent validation and improve the estimation accuracy of relatively small-size subgroups (e.g., cHCC-ICC, MIA, stage IB LUAD).

Conclusions

By integrating multiple fragmentomic features from cfDNA WGS data, our ensemble stacked model exhibited superior detection and localization power for the prevalent cancer types of PLC, CRC, and LUAD even at stages 0 and I. The robustness of our model is consistently high using as low as 1× sequencing coverage depth, suitable for developing accurate and affordable early detection assays for clinical practice.

Abbreviations

AUC: area under the curve; cfDNA: cell-free DNA; BPM: breakpoint motif; CI: confidence interval; CNV: copy number variation; ctDNA: circulating tumor DNA; CRC: colorectal adenocarcinoma; EDM: end motif; FSD: fragment size distribution; FSC: fragment size coverage; GBM: gradient boosting machine; GLM: generalized linear model; HCC: hepatocellular carcinoma; ICC: intrahepatic cholangiocarcinoma; LUAD: lung adenocarcinoma; MIA: minimally invasive adenocarcinoma; PLC: primary liver cancer; ROC: receiver operating characteristic; WGS: whole-genome sequencing.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12943-022-01594-w.

Additional file 1 : Supplementary Table 1. Participant demographics and baseline characteristics. **Supplementary Table 2**. Performance of the first-level cancer detection model. **Supplementary Table 3**. Performances of the cancer detection model on different subgroups based on clinical characteristics. **Supplementary Table 4**. Cancer detection model robustness test using downsampled (4× to 1× coverage depths) WGS data. Each downsampled coverage depth was repeated five times. **Supplementary Table 5**. Performances of the cancer detection model on extra healthy volunteer and at-risk patient cohort. **Supplementary Table 6**. The performance of the second level cancer origin model shown in the confusion matrix table. **Supplementary Table 7**. Cancer origin model robustness test using downsampled (4× to 1× coverage depths) WGS data. Each downsampled coverage depth was repeated five times.

Additional file 2: Supplementary Materials and Methods. Supplementary Figure 1. Distribution of cancer scores by cancer stages. **Supplementary Figure 2**. Cancer detection model evaluation using matched test cohort and cross-validated training cohort. **Supplementary Figure 3**. Evaluating cancer detection model sensitivity within different primary liver cancer subgroups. **Supplementary Figure 4**. Evaluating cancer detection model sensitivity within different colorectal cancer subgroups. **Supplementary Figure 5**. Evaluating cancer detection model sensitivity within different lung adenocarcinoma subgroups. **Supplementary Figure 6**. Cancer origin scores of the false-negative samples.

Acknowledgments

The authors thank all the patients, volunteers, and their families, the investigators, and the site personnel who participated in this study.

Authors' contributions

Sanjun C, SG, JP, JZ, and YS were responsible for the design of this work. ZW, XM, WG, Xiangyu Z, Xinyu W, YC, SM, NL, QM, YL, LZ, FT, QX, and FL enrolled patients and collected data. ZW, XM, WG, Xiangyu Z, Xinyu W, YC, SM, NL, QM, Xian Z, Hairong B, YJ, YL, LZ, FT, QX, SSY and FL elaborated the clinical data. YJ was responsible for the next-generation sequencing experiments. SW, XX, Shuang C, and YW analyzed the next-generation sequencing results. Hua B, ZW, XM, WG, Xiangyu Z, WT, and XC performed data analysis and interpretation. Hua B, SW, XX, Shuang C, YW, and RL contributed to machine learning modeling. Hua B, ZW, XM, WG, Xiangyu Z, WT, and XC wrote the manuscript. All authors contributed to manuscript editing and revision. All authors approved the submission of the final manuscript for publication.

Funding

This work was supported by the Institutional Fundamental Research Funds (2018PT32033 to SG), the CAMS Initiative for Innovative Medicine (2021-1-I2M-015 to SG), the Beijing Hope Run Special Fund of Cancer Foundation of China (LC2019B15 to WG), National Key R&D Program of China (2019YFC1315800 and 2019YFC1315802 to JZ), National Natural Science Foundation of China (81830102 and 81772578 to JZ, U1932145 to JP, 82002946 to YL, 82002451 to WG), Shanghai Municipal Key Clinical Specialty (to JZ), Science and Technology Commission of Shanghai Municipality (18401933402 to JP) and Shanghai Sailing Program (19YF1409500 to YL).

Availability of data and materials

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committees of the Chinese Academy of Medical Sciences, Peking Union Medical College Hospital, China-Japan Friendship Hospital, Zhongshan Hospital Fudan University, and Fudan University Shanghai Cancer Center, Shanghai Cancer Center

Institutional Review Board, and in accordance with international standards of good clinical practice. Written informed consent was obtained by all participants.

Consent for publication

We have received consent for publication from individuals who have participated in this study.

Competing interests

Hua B, WT, XC, SW, XX, Shuang C, YW, Xian Z, Hairong B, RL, SSY, YJ, Xue W, and YS are employees of Nanjing Geneseeq Technology Inc., China. All other authors have declared no conflicts of interest.

Author details

[1]Geneseeq Research Institute, Nanjing Geneseeq Technology Inc., Nanjing 210000, Jiangsu, China. [2]Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai 200032, China. [3]Key Laboratory of Carcinogenesis and Cancer Invasion (Fudan University), Ministry of Education, Shanghai 200032, China. [4]Shanghai Key Laboratory of Organ Transplantation, Zhongshan Hospital, Fudan University, Fudan University, 130 Fenglin Road, Shanghai 200032, China. [5]Department of Colorectal Surgery, Fudan University Shanghai Cancer Center, Shanghai 200032, China. [6]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China. [7]Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China. [8]Key Laboratory of Minimally Invasive Therapy Research for Lung Cancer, Chinese Academy of Medical Sciences, Beijing 100730, China. [9]Department of Thoracic Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing 100730, China. [10]Department of Thoracic Surgery, China-Japan Friendship Hospital, Beijing 100029, China. [11]Department of Cancer Institute, Fudan University Shanghai Cancer Center, Fudan University, Shanghai 200032, China. [12]Institute of Biomedical Sciences, Fudan University, Shanghai 200032, China. [13]State Key Laboratory of Genetic Engineering, Fudan University, Shanghai 200032, China. [14]School of Public Health, Nanjing Medical University, Nanjing 210029, Jiangsu, China.

Received: 1 February 2022 Accepted: 15 May 2022
Published online: 11 June 2022

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71:209–49.
2. World Health Organization. Guide to cancer early diagnosis: World Health Organization; 2017.
3. Chen X, Gole J, Gore A, He Q, Lu M, Min J, et al. Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. Nat Commun. 2020;11:3475.
4. Patel M, Shariff MI, Ladep NG, Thillainayagam AV, Thomas HC, Khan SA, et al. Hepatocellular carcinoma: diagnostics and screening. J Eval Clin Pract. 2012;18:335–42.
5. National Lung Screening Trial Research T, Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, et al. Results of initial low-dose computed tomographic screening for lung cancer. N Engl J Med. 2013;368:1980–91.
6. Daskalakis C, DiCarlo M, Hegarty S, Gudur A, Vernon SW, Myers RE. Predictors of overall and test-specific colorectal Cancer screening adherence. Prev Med. 2020;133:106022.
7. Stroun M, Maurice P, Vasioukhin V, Lyautey J, Lederrey C, Lefort F, et al. The origin and mechanism of circulating DNA. Ann N Y Acad Sci. 2000;906:161–8.
8. Sun K, Jiang P, Chan KC, Wong J, Cheng YK, Liang RH, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc Natl Acad Sci U S A. 2015;112:E5503–12.
9. Benesova L, Belsanova B, Suchanek S, Kopeckova M, Minarikova P, Lipska L, et al. Mutation-based detection and monitoring of cell-free tumor DNA in peripheral blood of cancer patients. Anal Biochem. 2013;433:227–34.
10. Lo YMD, Han DSC DSC, Jiang P, Chiu RWK. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. Sci. 2021;372:eaaw3616.
11. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. Nature. 2019;570:385–9.
12. Mathios D, Johansen JS, Cristiano S, Medina JE, Phallen J, Larsen KR, et al. Detection and characterization of lung cancer using cell-free DNA fragmentomes. Nat Commun. 2021;12:5060.
13. Liu J, Zhao H, Huang Y, Xu S, Zhou Y, Zhang W, et al. Genome-wide cell-free DNA methylation analyses improve accuracy of non-invasive diagnostic imaging for early-stage breast cancer. Mol Cancer. 2021;20:36.
14. Jiang P, Sun K, Tong YK, Cheng SH, Cheng THT, Heung MMS, et al. Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. Proc Natl Acad Sci U S A. 2018;115:E10925–33.
15. Chabon JJ, Hamilton EG, Kurtz DM, Esfahani MS, Moding EJ, Stehr H, et al. Integrating genomic features for non-invasive early lung cancer detection. Nature. 2020;580:245–51.
16. Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann Oncol. 2021;32:1167–77.
17. Chen L, Abou-Alfa GK, Zheng B, Liu JF, Bai J, Du LT, et al. Genome-scale profiling of circulating cell-free DNA signatures for early detection of hepatocellular carcinoma in cirrhotic patients. Cell Res. 2021;31:589–92.
18. Ma X, Chen Y, Tang W, Bao H, Mo S, Liu R, et al. Multi-dimensional fragmentomic assay for ultrasensitive early detection of colorectal advanced adenoma and adenocarcinoma. J Hematol Oncol. 2021;14:175.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC