

LETTER TO THE EDITOR

A serum microRNA signature for enhanced selection of people for lung cancer screening

Dear editor,

Lung cancer (LC) is the leading cause of cancer-related mortality globally [1]. Earlier detection by screening can substantially reduce LC mortality [2, 3], but should be focused on those at highest risk [4]. Risk stratification for LC screening, which is mostly based on smoking history [5], is far from perfect [6]. Therefore, additional criteria to better define those at highest risk of LC are needed to enhance the efficiency and cost-effectiveness of LC screening. Besides risk prediction models incorporating classical LC risk factors [7], blood-based biomarkers such as microRNAs (miRNAs) have emerged as potential candidates to improve LC risk prediction [8]. We aimed to derive and validate a blood-based miRNA signature predicting LC incidence in a large population-based cohort of older adults.

A two-stage study design with a marker discovery and a marker validation phase was applied (Supplementary Materials and Methods). In the discovery phase, plasma samples from 20 LC cases and 20 LC-free controls (discovery set) were profiled using next-generation sequencing (NGS), and 20 differentially expressed miRNA candidates were identified (Supplementary Table S1, Supplementary Figures S1 and S2). Additional candidates were selected from a previously conducted literature review using the following criteria: (1) miRNA evaluated in plasma or serum samples in Western populations; (2) miRNA included in a validated miRNA panel to discriminate LC cases from controls; (3) miRNA reported in ≥ 2 studies (Supplementary Table S2). In the marker validation phase, 40 miRNA candidates obtained through the NGS analyses and the lit-

erature review were measured by quantitative real-time polymerase chain reaction (qRT-PCR) in serum samples collected at baseline from participants of a prospective cohort of adults aged 50-75 years. The study population (prospective set, Supplementary Figure S3) included 237 incident LC cases and 191 randomly selected controls, i.e., participants without LC diagnosis until the end of 14 years of follow-up. Population characteristics are shown in Supplementary Table S3.

Analyses of the qRT-PCR data revealed that 20 of the 40 measured miRNAs were detectable in at least 90% of the samples that underwent qRT-PCR profiling (Supplementary Table S4), which were included in further analyses. To derive a multi-marker prediction algorithm that could effectively discriminate incident LC cases from LC-free controls, the least absolute shrinkage and selection operator (LASSO) regression model was applied to the 20 miRNAs based on the prospective set. Three miRNAs (miR-142-3p, miR-148a-3p and miR-451a) were selected. A continuous risk score, i.e., microRNA risk score (miR-score), was calculated for each participant in the prospective set using a linear combination of LASSO regression coefficient weighted expression values of the three miRNAs:

$$\text{miR - score} = -0.94121 + \text{miR - 142 - 3p} * -0.14913 + \text{miR - 148a - 3p} * 0.18115 + \text{miR - 451a} * 0.45143$$

Associations of the miR-score with LC incidence in prospective set participants were evaluated using logistic regression models adjusted for age, sex, smoking status and pack-years (Supplementary Tables S5 and S6). The miR-score was strongly associated with LC risk after adjustment for age, sex and smoking history comparing the highest quartile with the lower two quartiles (adjusted odds ratio [OR] = 5.01 [95% confidence interval = 2.94-8.69]). The OR per increase in the miR-score by 1 standard deviation was 2.23 (1.71-2.90) (Supplementary Table S6).

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; LASSO, least absolute shrinkage and selection operator; LC, lung cancer; LCRAT, Lung Cancer Risk Assessment Tool; LLP, Liverpool Lung Project Risk Model; LLPi, Liverpool Lung Project Incidence Risk Model; miRNA, microRNA; miR-score, microRNA risk score; NGS, next-generation sequencing; n, number; OR, odds ratio; PLCO_{M2012}, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Model 2012; qRT-PCR, quantitative real-time polymerase chain reaction.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Cancer Communications* published by John Wiley & Sons Australia, Ltd. on behalf of Sun Yat-sen University Cancer Center.

TABLE 1 Performance of LC models* [7] individually and in combination with the miR-score for LC risk prediction among ever-smokers ($n_{\text{cases/controls}} = 206/101$) in prospective set participants

LC model	AUC (95% CI)			P-value for improvement in AUC
	LC model only	LC model + miR-score apparent [#]	LC model + miR-score .632+ [§]	
LLPi	0.712 (0.651-0.774)	0.765 (0.708-0.821)	0.762	0.019
Pittsburgh Predictor	0.717 (0.656-0.779)	0.764 (0.707-0.821)	0.762	0.042
Bach	0.716 (0.654-0.778)	0.760 (0.703-0.817)	0.758	0.034
PLCO _{M2012}	0.694 (0.632-0.755)	0.757 (0.700-0.814)	0.754	0.008
LLP	0.669 (0.606-0.732)	0.750 (0.693-0.808)	0.747	0.003
Hoggart	0.655 (0.591-0.719)	0.738 (0.679-0.797)	0.734	0.004
Spitz	0.623 (557-0.689)	0.723 (0.665-0.781)	0.719	0.001
LCRAT	0.721 (0.658-0.783)	0.765 (0.707-0.822)	0.762	0.061

*R package “lcmmodels” (<https://dceg.cancer.gov/tools/risk-assessment/lcmmodels>) was used to calculate risks from the risk models.

[#]AUC not adjusted for overfitting

[§]AUC adjusted for overfitting by applying 0.632+ bootstrapping with 1000 replications.

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; LC, lung cancer; LCRAT, Lung Cancer Risk Assessment Tool; LLP, Liverpool Lung Project Risk Model; LLPi, Liverpool Lung Project Incidence Risk Model; miR-score, microRNA risk score; PLCO_{M2012}, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Model 2012.

Performance of pack-years individually and in combination with the miR-score for LC risk prediction in the prospective set was measured using the area under the receiver operating characteristic curves (AUCs). Potential over-optimism was accounted for by applying 0.632+ bootstrapping with 1000 replications. Performances of individual miRNAs from the miR-score for LC risk prediction in the prospective set were also measured. The optimism-corrected predictive accuracies of miR-142-3p, miR-148a-3p and miR-451a were 0.628, 0.679 and 0.691, respectively (data not shown). Compared to the performances of individual miRNAs, the miR-score exhibited higher predictive accuracy, yielding an optimism-corrected AUC of 0.740. Adding the miR-score to a model including pack-years improved the predictive performance (AUC: 0.754 vs. 0.808, $P < 0.001$), yielding an optimism-corrected AUC of 0.806 (Supplementary Table S7). Analyses stratified by age, sex, smoking history, and by time from blood sample collection to diagnosis consistently showed major increases in the AUC by adding the miR-score in all groups.

To compare the performance of the miR-score and combination of the miR-score and pack-years with the smoking-based criteria used in LC screening trials (Supplementary Table S8), we applied all these eligibility criteria to the prospective set. We observed that 77.5% (NLST criteria), 73.8% (MILD, DANTE, ITALUNG, DLCST criteria), 75.4% (LUSI, NELSON criteria) and 72.3% (DEPISCAN criteria) of LC-free participants were correctly identified as non-eligible and 54.9%, 59.9%, 57.8% and 60.3% of LC cases were correctly identified as eligible, respectively (Supplementary Table S9). At cutoffs identifying equal numbers of LC-free participants as non-eligible based on trial criteria, the corresponding proportions of LC cases identified

as eligible by the combination of the miR-score with pack-years were 65.4%, 73.4%, 72.2% and 76.4%, respectively, i.e., they were 10.5%, 13.5%, 14.4% and 16.1% higher than those obtained with the trial criteria; indicating that the miR-score combined with pack-years outperformed the criteria used in LC screening trials to identify target populations for LC screening.

In recent years, LC risk prediction models based on multiple LC risk factors have been proposed for risk stratification in LC screening [7]. Compared to the eligibility criteria used in the trials, the risk factor-based LC risk models have demonstrated a superior ability to select individuals for LC screening [7, 9, 10]. We evaluated the potential of improving LC risk prediction by combining established LC risk models with the miR-score within the risk group of ever-smokers among the prospective set. The predictive performances of eight individual LC risk models varied, with AUCs ranging from 0.623 to 0.721 (Table 1). The Lung Cancer Risk Assessment Tool (LCRAT) (AUC = 0.721, 95% CI = 0.658-0.783), the Pittsburgh Predictor (AUC = 0.717, 95% CI = 0.656-0.779) and the Bach models (AUC = 0.716, 95% CI = 0.654-0.778) performed best in predicting LC risk. Adding the miR-score to established LC risk models strongly increased the AUC between 0.041 and 0.096 (P values between 0.001 and 0.061). The highest optimism-corrected AUCs were observed for the combinations of the miR-score with the Liverpool Lung Project Risk Model (LLPi) (0.762), the Pittsburgh Predictor (0.762) and the LCRAT (0.762).

In conclusion, we derived and validated a miRNA signature that substantially improved LC risk prediction when combined with pack-years or risk factor-based LC risk models. Therefore, the miRNA signature could potentially

enhance LC risk stratification and might be useful for identifying high-risk populations for LC screening. Risk models incorporating the miR-score with established LC risk models could have important clinical implications for national LC screening programs and preventive strategies. Using blood samples routinely collected in medical practice, the miR-score could be calculated and employed in combination with established LC risk models for personalized LC risk prediction. Applying the miR-score could maximize screening benefits by facilitating early LC detection and prevention while minimizing the harms and costs associated with false-positive diagnoses. Individualized screening by risk models incorporating the miR-score could thus make low-dose computed tomography (LDCT)-based screening more efficient. External validation by prospective studies is required to verify the performance of our miRNA signature for LC risk stratification in different populations. Future research should address the acceptance, feasibility, and cost-effectiveness of such signatures in LC screening programs.

DECLARATIONS

ACKNOWLEDGMENTS

We gratefully acknowledge all the participants of the studies, the excellent cooperation and support of the participating partner practices and clinics, and the commitment of their physicians and staff who made this study possible. We thank the GEKKO Team at the National Center for Tumor Diseases (NCT) for outstanding engagement. GEKKO is conducted in the context of the NCT Early Detection and Prevention Programme (POC, NCT 3.0). We acknowledge the contributions of all the technicians and administrative staff in collecting and preparing the ESTHER biological samples.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

FUNDING INFORMATION

Haixin Yu was partly supported by the China Scholarship Council (No. 201606170104). The GEKKO study was funded by the German Federal Ministry of Education and Research, the German Cancer Aid, The German Cancer Research Center (GCRC/DKFZ), and the University Clinic Heidelberg (UKHD). There is no external sponsor. This funding source had no role in the study design, data collection, data analysis and interpretation, or the decision to submit results for presentation or publication. The ESTHER study was supported by the Baden-Württemberg State Ministry of Science, Research and Arts (Stuttgart, Germany); the Federal Ministry of Education and Research (Berlin, Germany); the Federal Ministry of Family Affairs, Senior Citizens, Women and Youth (Berlin, Germany);

and the Saarland state Ministry for Social Affairs, Health, Women and Family Affairs (Saarbrücken, Germany). The sponsors had no role in the study design, data collection, analysis and interpretation of data, and preparation, review, or approval of the manuscript.

AUTHORS' CONTRIBUTIONS

H.B. conceived and supervised the study. H.Y. and J.R.R. analyzed the data, interpreted the results and drafted the manuscript. B.S. and B.H. were responsible for coordinating the follow-up and work-up of follow-up data of the ESTHER study. M.B., Y.Z., E.S., P.S.-K, and H.B. critically reviewed the manuscript for important intellectual content and contributed to the revision of the manuscript. All authors reviewed and approved the final version for submission.

AVAILABILITY OF DATA AND MATERIALS

All data that support the findings of this study are available on reasonable request from the corresponding author (H.B.). The data are not publicly available because they contain information that could compromise research participant privacy/ consent.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The GEKKO study was approved by the ethics committees of the Medical Faculties of the University Heidelberg (S-392/2015), the Eberhard Karls University and the University Hospital Tübingen (876/2017BO2), the physicians' boards of Baden-Württemberg (B-F-2016-034) and Rhineland Palatinate (2018-13334_5). The ESTHER study was approved by the ethics committees of the Medical Faculty of the University of Heidelberg and the state medical board of Saarland, Germany. All participants provided written informed consent.

CONSENT TO PUBLICATION

Not applicable.

Haixin Yu^{1,2}
Janhavi R. Raut^{2,3} 
Megha Bhardwaj¹
Yan Zhang^{1,4}
Evelin Sandner⁵
Ben Schöttker^{1,6}
Bernd Hollecsek⁷
Petra Schrotz-King³
Hermann Brenner^{1,3,4}

¹Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg 69120, Baden-Württemberg, Germany

²Medical Faculty Heidelberg, University of Heidelberg, Heidelberg 69120, Baden-Württemberg, Germany

³Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg 69120, Baden-Württemberg, Germany

⁴German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg 69120, Baden-Württemberg, Germany

⁵Pneumology and Thoracic Oncology, Robert-Bosch Krankenhaus, Klinik Schillerhoehe, Gerlingen 70839, Baden-Württemberg, Germany

⁶Network Aging Research, University of Heidelberg, Bergheimer Straße 20, Heidelberg 69115, Baden-Württemberg, Germany

⁷Saarland Cancer Registry, Präsident-Baltz-Straße 5, Saarbrücken 66119, Saarland, Germany

Correspondence

Hermann Brenner, Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Im Neuenheimer Feld 581, D-69120 Heidelberg.
Email: h.brenner@dkfz-heidelberg.de

Haixin Yu and Janhavi R. Raut contributed equally to this work.

ORCID

Janhavi R. Raut  <https://orcid.org/0000-0001-6537-8598>

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(3):209–49.
2. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with

low-dose computed tomographic screening. *N Engl J Med.* 2011;365(5):395–409.

3. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N Engl J Med.* 2020;382(6):503–13.
4. Pinsky PF. Lung cancer screening with low-dose CT: a world-wide view. *Transl Lung Cancer Res.* 2018;7(3):234–42.
5. Krist AH, Davidson KW, Mangione CM, Barry MJ, Cabana M, Caughey AB, et al. Screening for Lung Cancer: US Preventive Services Task Force Recommendation Statement. *Jama.* 2021;325(10):962–70.
6. Rahal Z, El Nemr S, Sinjab A, Chami H, Tfayli A, Kadara H. Smoking and Lung Cancer: A Geo-Regional Perspective. *Front Oncol.* 2017;7:194.
7. Katki HA, Kovalchik SA, Petito LC, Cheung LC, Jacobs E, Jemal A, et al. Implications of Nine Risk Prediction Models for Selecting Ever-Smokers for Computed Tomography Lung Cancer Screening. *Ann Intern Med.* 2018;169(1):10–9.
8. Chu GCW, Lazare K, Sullivan F. Serum and blood based biomarkers for lung cancer screening: a systematic review. *BMC Cancer.* 2018;18(1):181.
9. Li K, Hüsing A, Sookthai D, Bergmann M, Boeing H, Becker N, et al. Selecting High-Risk Individuals for Lung Cancer Screening: A Prospective Evaluation of Existing Risk Models and Eligibility Criteria in the German EPIC Cohort. *Cancer Prev Res (Phila).* 2015;8(9):777–85.
10. Ten Haaf K, Jeon J, Tammemägi MC, Han SS, Kong CY, Plevritis SK, et al. Risk prediction models for selection of lung cancer screening candidates: A retrospective validation study. *PLoS Med.* 2017;14(4):e1002277.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.