Gut microbiota as non-invasive diagnostic and prognostic biomarkers for natural killer/T-cell lymphoma

We read with interest the study by Kartal *et al*[1] showing that the gut-microbiota-derived biomarkers for disease stratification are often shared by subjects across disease cohorts. Here, we confirmed their observations with findings from a newly diagnosed natural killer/T-cell lymphoma (NKTCL) cohort, in which the gut biomarkers were significantly overlapped with those of multiple disease cohorts and consistently enriched/depleted in subjects with those diseases. Importantly, many of the shared biomarkers were remarkably associated with patient outcomes in our cohort, implying that they may have broad prognostic values in multiple diseases.

'Microbiota-gut-lymphoma axis' represents a fascinating avenue of microbiota-mediated lymphomagenesis and intervention opportunity,[2] but the implications of gut microbiota in NKTCL remain enigmatic. To identify gut microbiota-derived diagnostic biomarkers for NKTCL, we recruited a discovery cohort consisting of 30 treatment-naïve patients and 20 healthy controls (HCs), and a validation cohort, including 12 patients and 13 HCs, respectively (online supplemental materials and methods). We applied shotgun metagenomic sequencing to their faecal samples, profiled their gut metagenomes using mOTUs2 V.2.5,[3] and trained a patient-stratification classifier with all species-level taxonomic features using the LASSO algorithm implemented in SIAMCAT.[4] Our classifier achieved an accuracy of 0.868 area under the receiver operating characteristic curve (AUROC) on the discovery cohort, and 0.910 AUROC on the validation cohort (figure 1A). To increase the sample size for model training, we retrained a LASSO classifier for the NKTCL using all the samples from both cohorts, and achieved an accuracy of 0.813 AUROC in cross-validation, which strongly support the role of gut microbiota as diagnostic biomarkers for NKTCL.

To examine the specificity of the NKTCL gut-microbiota-derived signature, we applied the all-sample NKTCL classifier to 29 public gut microbiota cohorts (online supplemental table S1). We observed an overall false positive rate (FPR) of 3.1% in the HCs, but higher FPRs in patients of several cohorts (figure 1B), especially those of the pancreatic cancer (Kartal_DE_2022_PC, Kartal_ES_2022_PC, Nagata_JP_2022_PC), Crohn's Disease (He_2017_CD, Franzosa_2018_CD, Forslund_2015_CD) and liver disease (Qin_2014_LD). These results imply significant overlaps in the biomarkers between these diseases and NKTCL, which was confirmed using LEfSe analysis[5] (figure 1C). Importantly, these biomarkers were consistently enriched/depleted in most cohorts, including the enrichment of oral-derived taxa of *Veillonella* and *Streptococcus* in the patients, and known beneficial species in HCs such as *Faecalibacterium prausnitzii*, *Eubacterium rectale* and *Bifidobacterium adolescentis*[1 6 7] (figure 1C). These findings indicate that our classifier can accurately distinguish NKTCL patients from HCs; nevertheless, due to the shared biomarkers with other diseases, combination of selected clinical indicators with microbial biomarkers would be salutary for a distinctive diagnostic model.

Survival data were available for the NKTCL patients in the discovery cohort. Notably, many identified microbiome biomarkers, especially those shared by multiple diseases, could significantly predict the overall survival (OS) and progression-free survival (PFS) of the patients, including *Streptococcus parasanguinis*, *Romboutsia timonensis* and *Veillonella atypica* (online supplemental figure 1A–D). Finally, we created a *Streptococcus parasanguinis–Romboutsia timonensis* index (SRI) as the relative abundance ratio of the two species, and obtained the best prognostic prediction power than other individual species and combinations. Namely, NKTCL patients with higher SRI scores showed significantly worse OS and PFS than those with lower SRI scores (figure 1D–E). Furthermore, we observed remarkable correlations between high SRI score and multiple adverse prognostic factors of NKTCL, including PINK-E, stage, lymph node

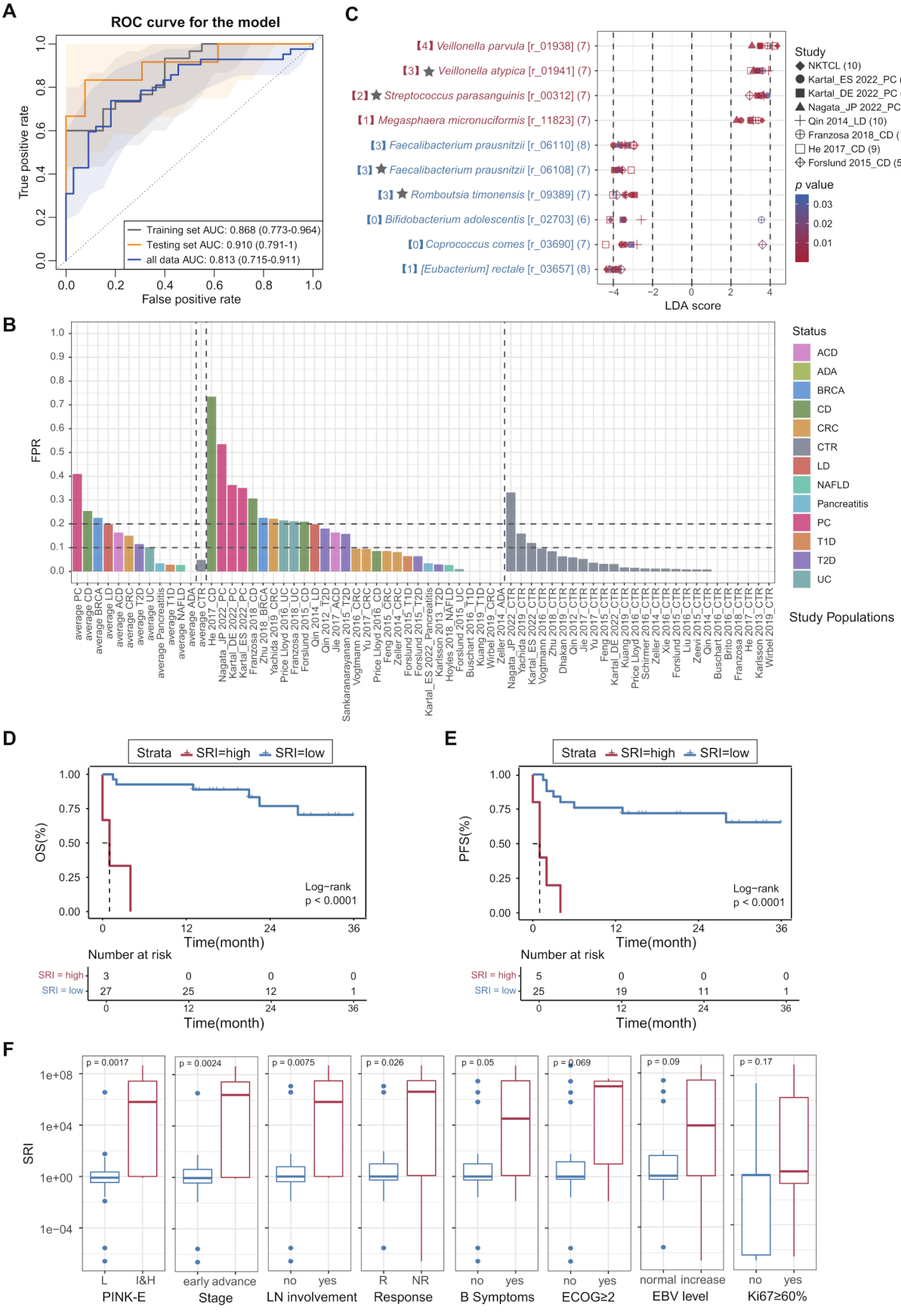

Figure 1 (A) Performance of the area under the receiver operating characteristic curve (AUROC) values of the gut microbiota-based classifier of NKTCL on the discovery cohort (threefold three times repeated cross-validation; grey line, the training set), the validation cohort (yellow line, the testing set), and all samples combined (ten-fold ten times repeated cross-validation; the 'all data model', blue line). (B) External validations of the

Figure 1 Continued

disease specificity of the NKTCL faecal microbiota model (the 'all data model'). False positive rates (FPRs) of the unconstrained model (without feature selection) in the 29 external test sets were shown as a bar plot. We defined the false-positive predictions as those wrongly classified as NKTCL by our model. Thus, two FPRs will be calculated for each cohort, one for the healthy controls (ie, the proportion of healthy controls that were wrongly classified as NKTCL), and another for the diseased individuals (ie, the proportion of diseased individuals that were wrongly classified as NKTCL). We then also calculated an overall FPRs for all the healthy controls and each of the diseases. Prediction results from the 'enrichment-constrained' model by selecting NKTCL-enriched biomarkers (enrichment-constrained model) as recommended by Kartal *et al*,[1] were shown in online supplemental figure 1E. (C) Marker microbes shared by the NKTCL cohort and other seven cohorts that had ~20% and higher FPRs in their diseased subjects in (B); markers were identified using the LDA Effect Size (LEfSe) analysis. Red (blue) species name represents its enriched (depleted) in patients. Wilcoxon rank sum test was used to compare the differences in relative abundances between the patients and HCs of the respective cohorts. Inside the square brackets are the numbers of studies in which the species were also among the top features (robustness >50%) of the corresponding disease-stratification classifiers (online supplemental table S2). The 'Star' symbol in front of a species name indicates that the species are significantly associated with patients' survival in our NKTCL cohort; the details can be found in online supplemental figure 1A–D. Inside the parentheses next to the species name is the number of studies in which the corresponding species were identified as a biomarker, that is, with |LDA|≥2. Inside the parentheses after a study name is the total number of species in this figure that were also biomarkers of the study. (D–E) the overall survival (OS) and progression-free survival (PFS) Kaplan-Meier survival curves for NKTCL patients (n=30). Patients were divided into the SRI-high group and SRI-low group according to scores of the *Streptococcus parasanguinis–Romboutsia timonensis* index (SRI), calculated using the quotient of the relative abundances of the two species; the cut-points of SRI 26386550 for OS and 10 776 890 for PFS, and were determined by the 'survminer' R package V.0.4.9[8] (https://github.com/kassambara/survminer). Log-rank test was used to calculate the p values. (F) Correlations between the SRI score and multiple adverse prognostic factors of NKTCL, including prognostic index for natural killer lymphoma-Epstein-Barr virus (PINK-E; L: low risk, I: intermediate risk, H: high risk), disease stage, lymph node (LN) involvement, responses to first-line treatment (R: response, NR: non-response), B symptoms, Eastern Cooperative Oncology Group (ECOG) Performance Status ≥ 2, an increase in plasm Epstein-Barr virus (EBV) DNA level, and Ki67 expression ≥60%. Wilcoxon rank sum test was used to compare continuous variables between groups. (More specific descriptions on these results could be found in online supplemental results). ACD, atherosclerotic coronary disease; ADA, American diabetes; BRCA, breast cancer; CD, Crohn's disease; CRC, colorectal cancer; CTR, controls; DE, German; ES, Spanish; JP, Japan; LD, liver disease; NAFLD, non-alcoholic fatty liver disease; PC, pancreatic cancer; T1D, type 1 diabetes; T2D, type 2 diabetes; UC, ulcerative colitis.

involvement and responses to first-line treatment (all $p<0.05$; figure 1F).

Overall, our results lend support for gut microbiota as a potent assistive diagnostic tool for NKTCL. Moreover, the SRI score, based on the shared biomarkers, may have extensive prognostic utility in multiple diseases and deserves further scrutiny (online supplemental discussion).

Zhuangzhuang Shi,[1] Guoru Hu,[2] Min W Li,[2] Lei Zhang,[1,3] Xin Li,[1,3] Ling Li,[1,3] Xinhua Wang,[1,3] Xiaorui Fu,[1,3] Zhenchang Sun,[1,3] Xudong Zhang,[1,3] Li Tian,[1,3] Zhaoming Li,[1,3,4,5] Wei-Hua Chen ,[2,6,7] Mingzhi Zhang[1,3,4]

[1]Department of Oncology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China
[2]Department of Bioinformatics and Systems Biology, Huazhong University of Science and Technology College of Life Sciences and Technology, Wuhan, Hubei, China
[3]Lymphoma Diagnosis and Treatment Centre of Henan Province, Zhengzhou, Henan, China
[4]State Key Laboratory of Esophageal Cancer Prevention & Treatment and Henan Key Laboratory for Esophageal Cancer Research, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China
[5]Academy of Medical Sciences of Zhengzhou University, Zhengzhou University, Zhengzhou, Henan, China
[6]Institution of Medical Artificial Intelligence, Binzhou Medical University, Yantai, Shandong, China
[7]College of Life Science, Henan Normal University, Xinxiang, Henan, China

Correspondence to Professor Mingzhi Zhang and Professor Zhaoming Li, Department of Oncology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China; mingzhi_zhang1@163.com, fcclizm@zzu.edu.cn and Professor Wei-Hua Chen, Department of Bioinformatics and Systems Biology, Huazhong University of Science and Technology College of Life Sciences and Technology, Wuhan, Hubei, China; weihuachen@hust.edu.cn

Acknowledgements We would like to thank all the clinical doctors from the Lymphoma Diagnosis and Treatment Centre of Henan Province for their kind suggestions, and we also thank all the generous participants of this study for their supports.

Contributors Study concept and design: MZ, W-HC and ZL. Samples collection: ZS, LZ, XL, XW, LL and XF. Data acquisition: ZS, GH, MWL, ZS, ZL, XZ and LT. Analysis and interpretation of data: W-HC, MZ, ZL, ZS, GH and MWL. Technical and material support: LZ, XL, XW, LL, XF, ZS, ZL, XZ and LT. Drafting of the manuscript: GH and ZS. Revising of the manuscript: W-HC, ZL and MZ. All the authors approved the final version of the manuscript.

Funding This work was supported by the National Natural Science Foundation of China (81970184; 82170183; U1904139; 82070209; 82070210).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study was performed in accordance with the Declaration of Helsinki and rules of good clinical practice, and the study was approved by the Ethics Review Committee of the First Affiliated Hospital of Zhengzhou University, Zhenzhou, China. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2022-328256

►

►).

ZS and GH contributed equally.

ZS and GH are joint first authors.

Check for updates

To cite Shi Z, Hu G, Li MW, *et al*. *Gut* 2023;**72**:1999–2002.

Received 11 July 2022
Accepted 30 September 2022
Published Online First 8 November 2022

Gut 2023;**72**:1999–2002. doi:10.1136/gutjnl-2022-328256

ORCID iD
Wei-Hua Chen http://orcid.org/0000-0001-5160-4398

REFERENCES

1 Kartal E, Schmidt TSB, Molina-Montes E, *et al*. A faecal microbiota signature with high specificity for pancreatic cancer. *Gut* 2022;71:1359–72.
2 Shi Z, Zhang M. Emerging roles for the gut microbiome in lymphoid neoplasms. *Clin Med Insights Oncol* 2021;15:117955492110241.
3 Milanese A, Mende DR, Paoli L, *et al*. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat Commun* 2019;10:1014.
4 Wirbel J, Zych K, Essex M, *et al*. Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. *Genome Biol* 2021;22:93.
5 Segata N, Izard J, Waldron L, *et al*. Metagenomic biomarker discovery and explanation. *Genome Biol* 2011;12:R60.
6 Montalban-Arques A, Katkeviciute E, Busenhart P, *et al*. Commensal Clostridiales strains mediate effective anti-cancer immune response against solid tumors. *Cell Host Microbe* 2021;29:1573–88.
7 Nagata N, Nishijima S, Kojima Y, *et al*. Metagenomic identification of microbial signatures predicting pancreatic cancer from a multinational study. *Gastroenterology* 2022;163:222–38.
8 Kassambara AKM, Biecek P, Fabian S. *Survminer: Drawing survival curves using 'ggplot2'*, 2021.