

RESEARCH

Open Access



A molecular approach integrating genomic and DNA methylation profiling for tissue of origin identification in lung-specific cancer of unknown primary

Kaiyan Chen^{1,2,3}, Fanrong Zhang^{1,2,4}, Xiaoqing Yu^{1,2,5}, Zhiyu Huang^{1,2,3}, Lei Gong^{1,2,3}, Yanjun Xu^{1,2,3}, Hui Li^{1,2,3}, Sizhe Yu^{1,2,3} and Yun Fan^{1,2,3,6*}

Abstract

Background: Determining the tissue of origin (TOO) is essential for managing cancer of unknown primary (CUP). In this study, we evaluated the concordance between genome profiling and DNA methylation analysis in determining TOO for lung-specific CUP and assessed their performance by comparing the clinical responses and survival outcomes of patients predicted with multiple primary or with metastatic cancer.

Methods: We started by retrospectively screening for CUP patients who presented with both intra- and extrathoracic tumors. Tumor samples from included patients were analyzed with targeted sequencing with a 520-gene panel and targeted bisulfite sequencing. TOO inferences were made in parallel via an algorithm using genome profiles and time interval between tumors and via machine learning-based classification of DNA methylation profiles.

Results: Four hundred patients were screened retrospectively. Excluding patients definitively diagnosed with conventional diagnostic work-up or without available samples, 16 CUP patients were included. Both molecular approaches alone enabled inference of clonality for all analyzed patients. Genome profile enabled TOO inference for 43.8% (7/16) patients, and the percentage rose to 68.8% (11/16) after considering inter-tumor time lag. On the other hand, DNA methylation analysis was conclusive for TOO prediction for 100% (14/14) patients with available samples. The two approaches gave 100% (9/9) concordant inferences regarding clonality and TOO identity. Moreover, patients predicted with metastatic disease showed significantly shorter overall survival than those with multiple primary tumors.

Conclusions: Genome and DNA methylation profiling have shown promise as individual analysis for TOO identification. This study demonstrated the feasibility of incorporating the two methods and proposes an integrative scheme to facilitate diagnosing and treating lung-specific CUPs.

Keywords: Cancer with unknown primary (CUP), Multiple primary tumor, Comprehensive genomic profiling (CGP), DNA methylation, Machine learning, Lung cancer

Introduction

Lung cancer (LC) is the second most frequent cancer worldwide [1]. Approximately 50% of lung cancer patients present with distant metastases at diagnosis [2]. Meanwhile, the lung is a common destination for

*Correspondence: fanyun@zjcc.org.cn

⁶ Department of Thoracic Medical Oncology, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou 310022, 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.

metastatic spread from other primary malignancies such as colorectal, cervical, and gastric cancer [3–6]. For patients presenting with intra- and extrathoracic tumors, it is paramount to elucidate the tumor relationship due to the therapeutic implications. Despite advances in conventional diagnostic work-up, histopathologically indeterminate cases still arise, constituting cancers of unknown primary (CUPs). CUP are typically characterized by dismal prognosis, with median overall survival of 1 year [7]. There is an unmet need for tissue of origin (TOO) identification for patients with lung-specific cancer of unknown primary (CUP).

A handful of studies have explored the value of different molecular tests for TOO identification [8–10]. Notably, gene expression profiles (GEPs) were used in a phase II trial that randomized CUP patients to standard chemotherapy and site-specific therapy guided by TOO inferred from GEP [11]. A later trial combined site-specific treatment based on GEPs and targeted therapy guided by genome profiles [12]. In both investigations, median overall survival (OS) and progression-free survival (PFS) were better for those predicted with more-responsive than with less-responsive tumor types, although no significant difference was observed for OS, PFS, and 1-year survival between site-specific and standard chemotherapy [11, 12]. In addition, there have been studies that used genomic alterations to distinguish multiple primary LC from intrapulmonary metastasis [13, 14] or DNA methylation to classify tumor type [15–17]. As a major epigenetic mechanism, DNA methylation refers to methylation generally of the cytosine base within the context of CpG dinucleotide repeats [18]. Features of DNA methylation patterns, such as hyper- or hypomethylation at certain sites and the relative ordering of relative methylation orderings of CpG sites, show marked tissue specificity [16, 19]. Despite encouraging advances, so far most investigations on TOO determination have focused on individual molecular analysis. Therefore, TOO identification and treatment of CUP patients may benefit from integration of different molecular tests.

In this study, we aim to establish an integrative approach for TOO inference based on multidisciplinary evidence. To this end, we retrospectively 16 patients with lungs-specific CUPs. The tumor origins were inferred from genome profiles and clinical information and from DNA methylation profiles, respectively. We then evaluated the level of concordance between these two methods and whether the predicted diagnoses were consistent with clinical outcomes. Furthermore, we propose an integrative molecular approach based on these results for TOO identification.

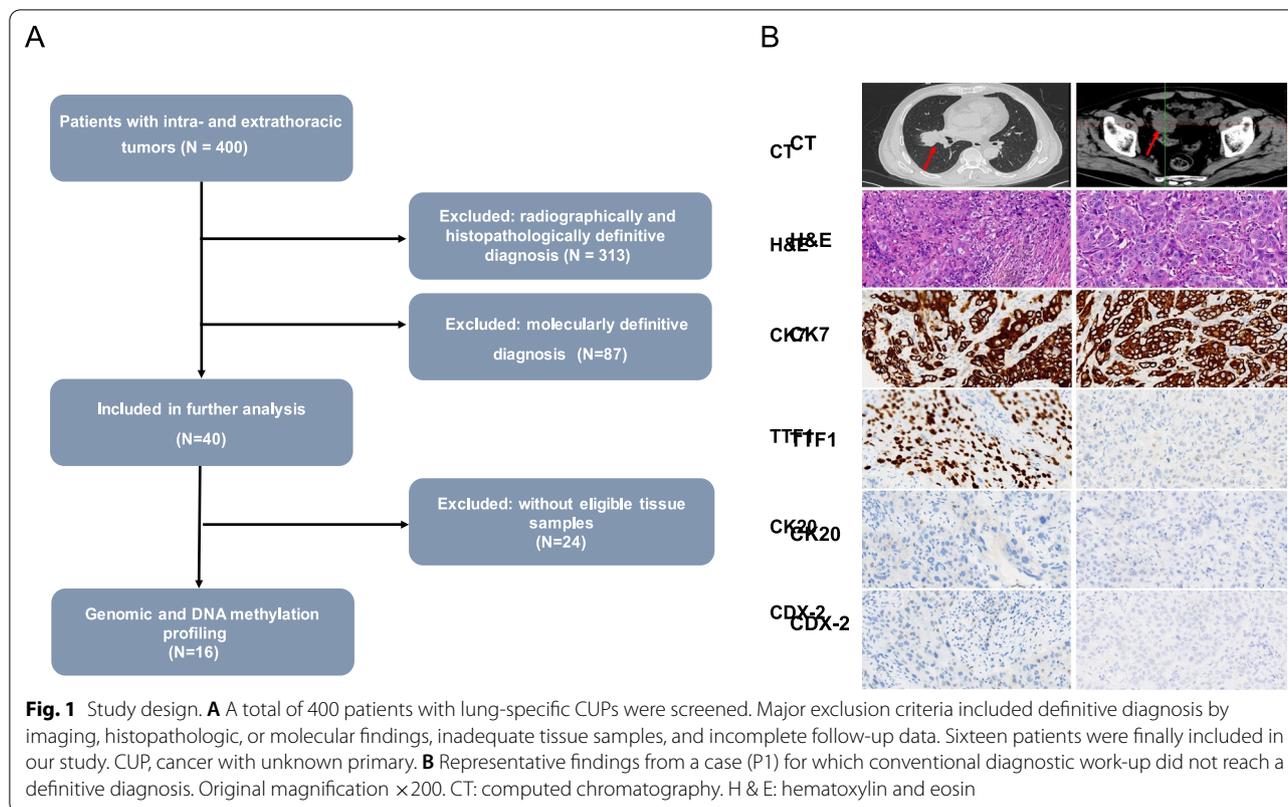
Methods

Patients

We retrospectively screened for 400 patients who presented with intrathoracic and extrathoracic tumors and visited Zhejiang Cancer Hospital between March 2019 and January 2021. Extrathoracic tumors were defined as lesions other than those in the lungs and corresponding regional lymph nodes per the eighth edition of the tumor-node-metastasis (TNM) classification of lung cancer [20]. Eligibility criteria included a diagnosis of lung-specific CUP after comprehensive examination, including pathological evaluation by immunohistochemistry, chest-abdomen-pelvis computed tomography scans, and directed assessment of all symptomatic areas. Results of the laboratory tests, histopathologic work-up, and molecular and imaging studies were assessed by an institutional multidisciplinary team. Consensus regarding tumor clonality and origin was reached in 360 of 400 patients. As the flow chart in Fig. 1A indicates, excluding 24 patients with no eligible samples, 16 patients were finally included in this study. Differential diagnosis for all 16 patients was conducted using immunochemical staining with tumor type-specific markers, including CK20 and CDX-2 for colorectal, TTF1 and NapsinA for pulmonary, CK7 and CK20 for gastric cancer, and P16 for cervical cancer. Collection and analysis of tumor samples were approved by the Institutional Review Board of Zhejiang Cancer Hospital (No.IRB-2021-54). Informed consent was obtained from all patients.

Comprehensive genomic profiling (CGP) analysis and inference of tumor relationship

For CGP, all 16 included patients provided one samples for the intrathoracic tumor and one for the extrathoracic tumors. A total 32 formalin-fixed and paraffin-embedded (FFPE) tissue specimens were subjected to targeted sequencing with a panel of 520 cancer-related genes (OncoScreen Plus, Burning Rock Biotech, Guangzhou, China). A sizable number of clinical investigations in multiple tumor types have conducted molecular testing using this panel, including lung cancer, colorectal cancer, and cervical cancer [6, 21, 22]. DNA extraction, quantification, library construction, sequencing, and data processing were performed as previously described [23]. A more detailed description is provided in the Additional file 1. The sequencing analysis detects single nucleotide variants (SNVs), small insertions and deletions, rearrangements, and splice variants. Tumor relationship was inferred via a multi-step algorithm. Multiple primary tumors were predicted if the tumor samples shared no genomic alteration and metastatic cancer was predicted if the tumors shared clinically actionable *EGFR* mutations or deletions or *ALK* or *ROS1* rearrangements.



The remaining cases were assigned as “inconclusive as to tumor relationship” if they met neither of the following criteria: (i) both tumors harbored ≥ 5 genomic alterations and ≥ 2 of which were shared, or (ii) either tumor harbored < 5 alterations and ≥ 1 of which were shared. If any of these criteria was met, we further considered the time interval between manifestation of the tumors. For metachronous cases (detected at an interval of ≥ 6 months), the inferred diagnosis was metastatic disease originating from the organ involved in the earlier tumor; otherwise the case would be regarded as metastatic cancer but inconclusive as to TOO.

DNA methylation profiling

A total of 29 samples were available for analysis, including paired samples of intra- and extrathoracic tumors from 14 patients and a sample of gastric tumor from one patient. DNA methylation profiling as performed as previously described [24]. Briefly, the bisulfite sequencing (BS-seq) library was prepared with the brELSATM method. Custom-designed methylation profiling RNA baits were used for target enrichment which covers 80,672 CpG sites and spans 1.05 mega base of human genome. The target libraries were quantified and sequenced on NovaSeq 6000 (Illumina, San Diego, CA, USA) with an average sequencing depth of $1000\times$.

Further bioinformatic analyses were performed to remove custom adaptor sequences and low-quality bases, align and merge paired-end reads, and build methylation blocks. A more detailed description is provided in the Additional file 1.

Construction of tissue classification models and TOO inference based on DNA methylation profiles

DNA methylation-based classifiers were constructed using a previously described machine learning approach [19]. Briefly, differential methylation sites were selected from The Cancer Genome Atlas database and further segregated into blocks, and the “block-level” methylation profile was represented with a matrix named “MBS”. A support vector machine classifier was implemented to construct classification models that categorized the tumor binarily as LC-origin or non-LC-origin. The algorithm maps training samples to points in a high-dimensional space. The width of the gap between two categories was maximized, and each mis-mapped training sample was penalized according to the parameter set to the model. A proprietary dataset was used for model training, which consisted of 70 tumor tissue samples from four sites, including 22 from lungs, 19 from stomach, 16 from colon, and 13 from cervix (Burning Rock Biotech, Guangzhou, China). Three classifications models

were constructed to assign the tumor origin as the lung or one of the three alternative organs. In each model, LC and non-LC training samples were labeled with 0 and 1, respectively. The desired parameter was achieved through fivefold cross-validation using the training samples.

For TOO inference, DNA methylation profiles from the 29 samples were subjected to one of the three models based on the site of the extrathoracic tumor. These profiles were mapped into the same high-dimensional space in the identification models and assigned a Methyl Score based on the detected TOO-related methylation signals. As determined via the Youden index, the best cutoff (0.5) was applied to distinguish between LC-origin (Methyl Score < 0.5) and non-LC-origin (Methyl Score > 0.5) tumor tissue samples.

Assessment of clinical response

Patients were evaluated for responses every 6 weeks after treatment onset until the detection of tumor progression or treatment completion, at which point they were all evaluated for survival outcomes. Tumor responses were assessed according to Response Evaluation Criteria in Solid Tumors, version 1.1.

Statistical analyses

Statistical analyses were performed with the statistical programming language R and GraphPad Prism. Survival curves were estimated with the Kaplan–Meier approach and tested for significance using the log-rank test. Statistical significance was defined as $P < 0.05$ in a two-sided test.

Results

Patient characteristics

In this single-center retrospective study, we identified 400 patients with LC and a synchronous or metachronous tumor involving another organ. Despite thorough analysis and consultation with a multidisciplinary tumor board, 40 cases (10.0%) remained indeterminate due to inconclusive clinical, histopathologic, and imaging findings. Excluding 24 individuals without eligible tissue samples, matched tissues from 16 cases were analyzed for genomic aberrations and/or DNA methylation profiles (Fig. 1A). Figure 1B shows representative findings from a case for which multidisciplinary consulting did not reach a definitive diagnosis based on conventional diagnostic work-up. Clinicohistologic characteristics of these 16 patients are summarized in Table 1. The median age was 53 years, with a range of 47–77 years. Among them, 62.5% ($n = 10$) were male, 43.8% ($n = 7$) had a family history of cancer, and 31.3% ($n = 5$) were ever-smokers. Most patients had a good ECOG performance status of 1

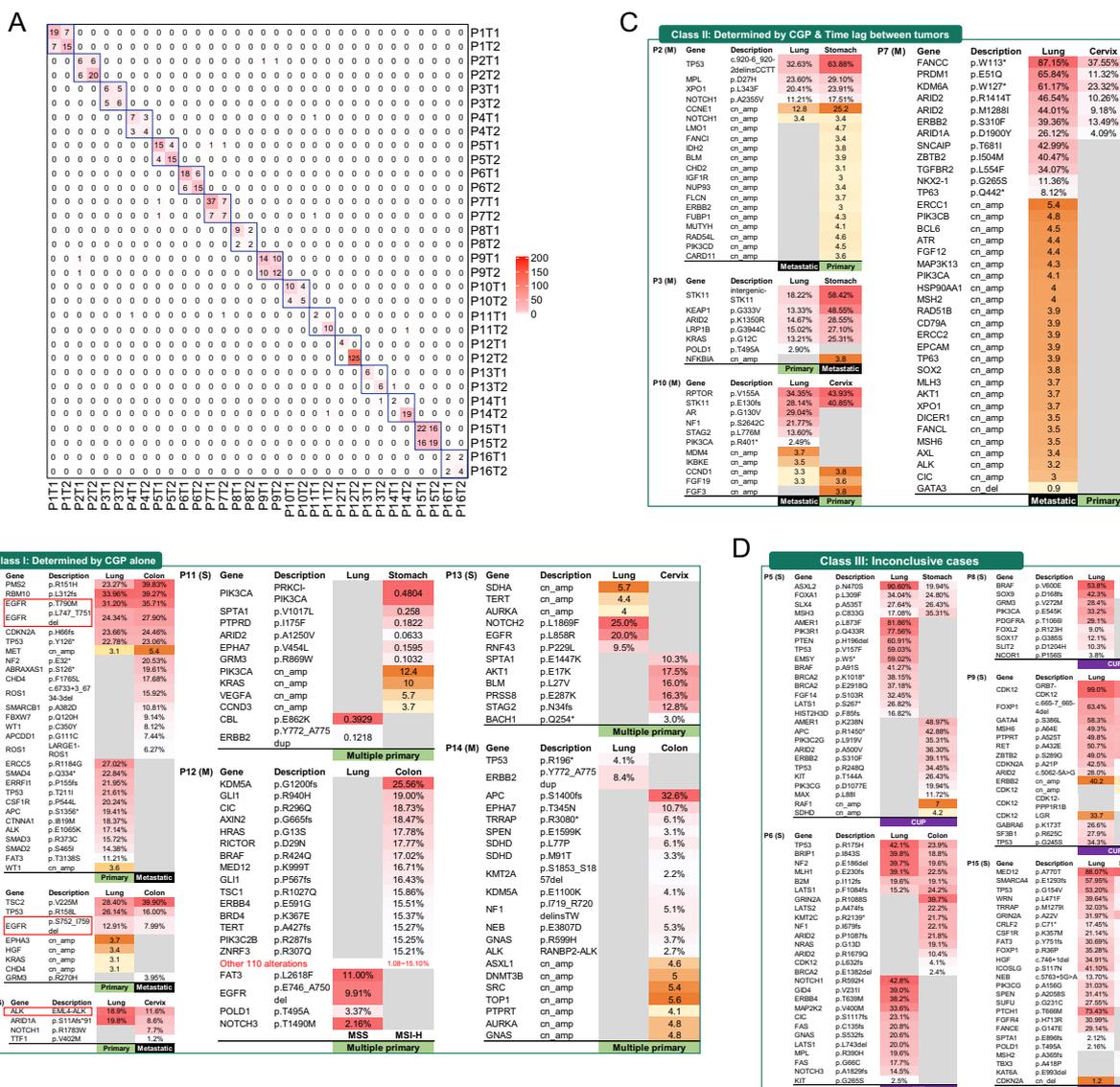
Table 1 Baseline clinicopathologic characteristics of the patients included in this study

Characteristic	No. (%)
Age, median (range), y	53 (47–77)
Sex	
Male	10 (62.5)
Female	6 (37.5)
ECOG performance status	
1	12 (75.0)
2	4 (25.0)
Extrathoracic lesions	
Stomach	5 (31.2)
Colon/rectum	7 (43.8)
Cervix	4 (25.0)
Histology	
Adenocarcinoma	11 (68.8)
Squamous cell carcinoma	1 (6.3)
Poorly-undifferentiated carcinoma	3 (18.8)
Other	1 (6.3)
Smoking history	
Never-smoker	11 (68.6)
Ever-smoker	5 (31.3)
Family history of cancer	
No	9 (56.3)
Yes	7 (43.8)

($n = 12$, 75.0%). The extrathoracic tumors were located in the stomach ($n = 5$), intestine ($n = 7$) and cervix ($n = 4$). According to the WHO classification for tumors, 68.8% ($n = 11$) of cases were identified as adenocarcinoma. The remaining five (31.2%) non-adenocarcinoma cases consisted of three poorly-undifferentiated carcinomas, one squamous carcinoma, and one atypical carcinoid.

TOO inference based on comprehensive genomic profiling (CGP)

CGP was performed on matched intra- and extrathoracic tumor tissues from all included 16 patients, followed by inference of tumor clonality and origin using a multi-step algorithm. As summarized in Fig. 2A, no alteration was shared between the matched samples for patients 11–14 despite the relatively large number of detected alterations (range of sum, 12–125). Detailed identities and abundances of these genomic aberrations are listed in Fig. 2B–D. Based on the presence of clinically relevant, high-specific driver alterations in LC, such as *EGFR* exon 19 deletion and *EML4-ALK* rearrangement, metastatic cancer with an LC origin was inferred for the corresponding patients 1, 4, and 16. As such, CGP data alone enabled unambiguous TOO identification for these 7 patients, who were grouped as Class I patients



detected from a training set of 70 tumor samples of known origins [19]. As tissue specificity has been established in DNA methylation patterns and serves as the basis for TOO identification [25], we tested the validity of our data by interrogating the distinctiveness of profiles from different organs. Indeed, principal component analysis indicated separate or largely non-overlapping clusters consisting mainly of samples from the same disease sites (Fig. 3A–C). Cross-validation suggested high performance of all three models in distinguishing tumors of LC versus non-LC origin, as indicated by the high areas under the curves (0.99 for all; Fig. 3D, F, Additional file 1: Figs. S1, and S2).

Of the 16 included patients, 14 had 2 DNA methylation profiles each, 1 patient (P11) had 1 available profile and 1 patient (P16) had none (Table 2). A total of 29 profiles were therefore which subjected to a site-appropriate classification model. As summarized in Table 2, DNA methylation analysis was able to unequivocally assign TOO for all 14 patients with methylation profiles from both tumors. Moreover, of the 9 patients with TOO inferred by both CGP-time lag and DNA methylation analyses, all had identical inferences. For P11 and P16, CGP was able to unambiguously assign TOOs for their tumor samples and matched the DNA methylation-based inference for the one sample from P11. In summary, DNA methylation profiling achieved a higher inference rate (100%) than

the CGP-time lag approach (68.8%) and 100% agreement with latter in terms of TOO identity (9/9).

Survival outcomes of patients with lung-specific CUPs

By-patient details of treatment regimens and survival outcomes are provided in Additional file 2: Table S1. After a median follow-up time of 67.4 months, the median OS was 52.6 months (95% confidence interval 33.1–72.1; Fig. 4A), which exceeded those reported in most studies of CUPs [11, 12]. As expected, patients diagnosed with metastatic disease had a significantly shorter OS (median 51.5 months) than those with multiple primary tumors (median 74.1 months; log-rank test $p=0.028$; Fig. 4B), which attested to the accuracy of the CGP-time lag and DNA methylation analyses. TOO determination has significant therapeutic implications, which can be illustrated by the courses of management for patients 1 and 5. P1 was highlighted in Fig. 1B as a representative CUP case. The actual diagnosis coincided with our inference of lung cancer with colon metastasis (Table 2). The patient was treated with an EGFR inhibitor as first-line therapy based on axillary nodal disease and detection of activating *EGFR* mutations (Fig. 2B). After a PFS of 8 months and a best response of partial response (PR), he presented with bowel bleeding due to disease progression and underwent salvage surgery to remove the rectal lesion. Pemetrexed and carboplatin

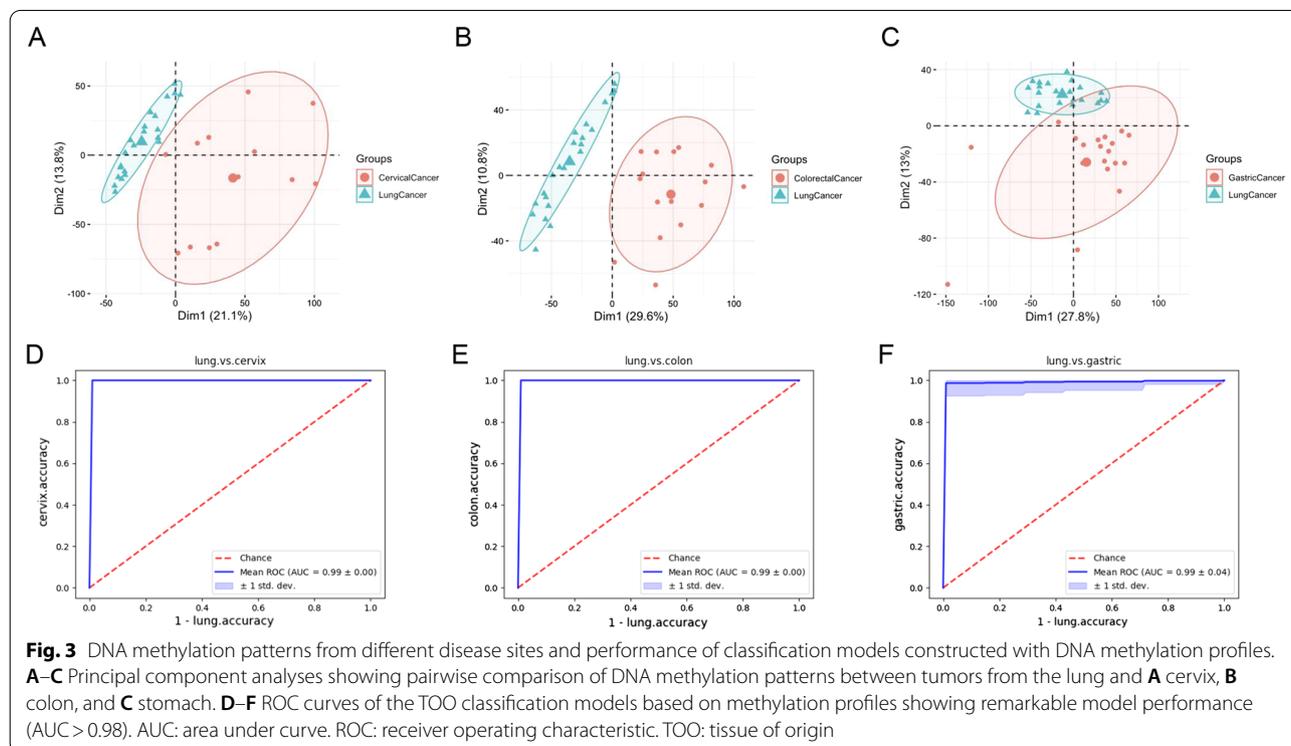


Table 2 Inferred tumor relatedness and origin for the 16 patients with lung-specific cancer of unknown primary

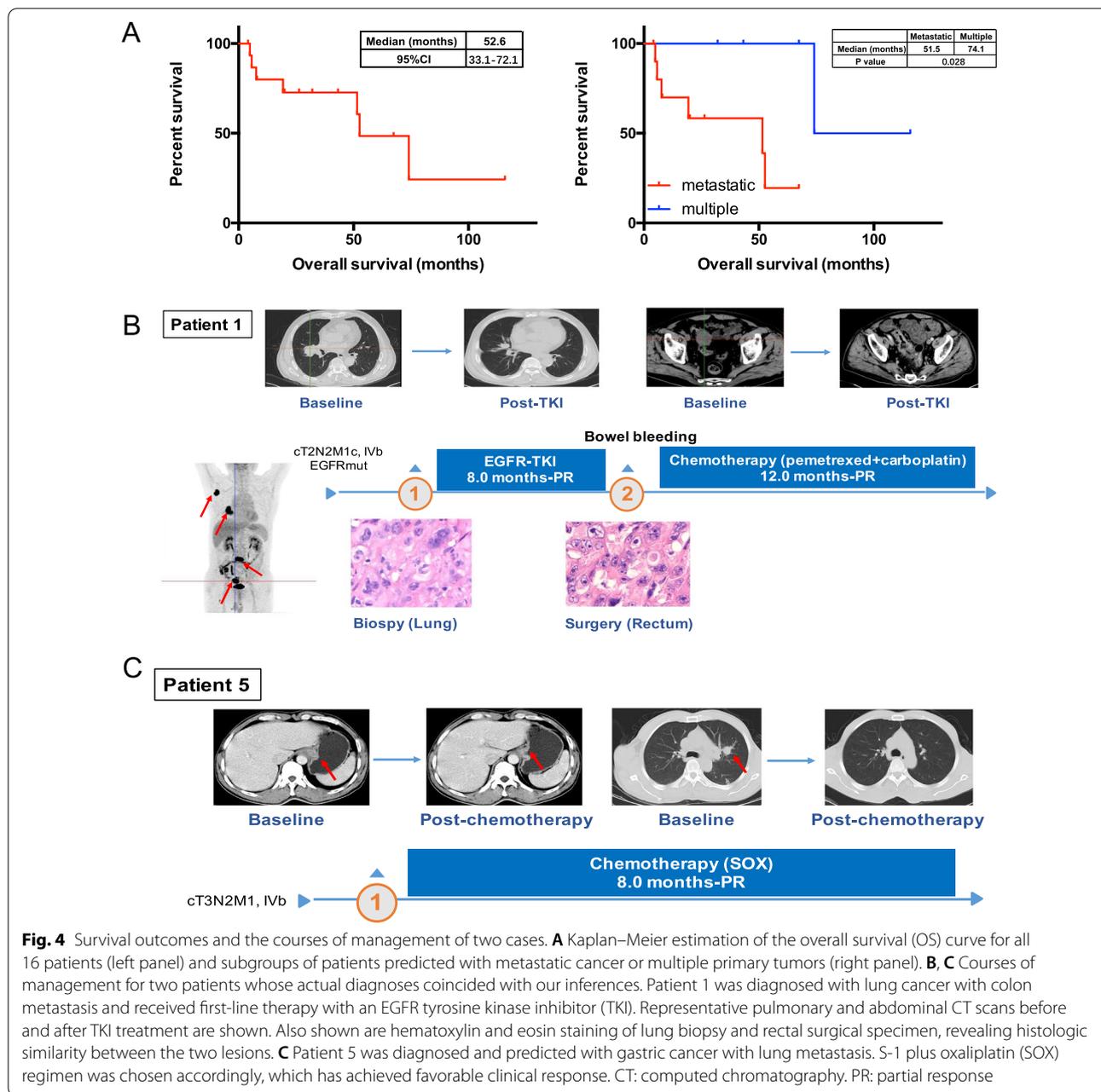
Patient ID	Time lag between tumors	Tumor location	Primary vs. metastatic lesions			CGP-methylation concordance
			Per CGP analysis	Per DNA methylation analysis	Methyl score	
1	Synchronous	Lung	P	P	0.475	Y
		Colon	M	M	0.378	
2	Metachronous	Lung	M	M	0.537	Y
		Stomach	P	P	0.634	
3	Metachronous	Lung	P	P	0.393	Y
		Stomach	M	M	0.354	
4	Synchronous	Lung	P	P	0.349	Y
		Colon	M	M	0.391	
7	Metachronous	Lung	M	M	0.695	Y
		Cervix	P	P	0.666	
10	Metachronous	Lung	M	M	0.715	Y
		Cervix	P	P	0.713	
12	Metachronous	Lung	P	P	0.259	Y
		Colon	P	P	0.647	
13	Synchronous	Lung	P	P	0.289	Y
		Cervix	P	P	0.548	
14	Metachronous	Lung	P	P	0.249	Y
		Colon	P	P	0.797	
5	Synchronous	Lung	Inconclusive	M	0.531	/
		Stomach		P	0.720	
6	Synchronous	Lung	Inconclusive	P	0.297	/
		Colon		M	0.652	
8	Synchronous	Lung	Inconclusive	M	0.575	/
		Colon		P	0.554	
9	Synchronous	Lung	Inconclusive	P	0.491	/
		Colon		M	0.455	
15	Synchronous	Lung	Inconclusive	P	0.428	/
		Stomach		P	0.556	
11	Synchronous	Lung	P	Not available	/	Y
		Stomach	P	P	0.751	
16	Synchronous	Lung	P	Not available	/	/
		Cervix	M	Not available	/	

wad then administered as second-line therapy due to its well-established efficacy in lung cancer was then administered as second-line treatment [26], which achieved a PFS of 12 months as of the latest follow-up. Had colon been identified as the tumor origin, the patient could have received a chemotherapeutic regimen for instead of EGFR inhibitor as first-line treatment. P5 was another case in which the actual and putative diagnoses agreed. As shown in Fig. 4C, the patient was diagnosed with gastric cancer with pulmonary metastasis and treated with S-1 plus oxaliplatin, to which he responded favorably with a best response of PR and a PFS of 8 months.

Together, the favorable clinical response, along with the markedly different survival outcomes between putative metastatic and multiple primary patients, suggested promising clinical value of the two molecular analyses.

Discussion

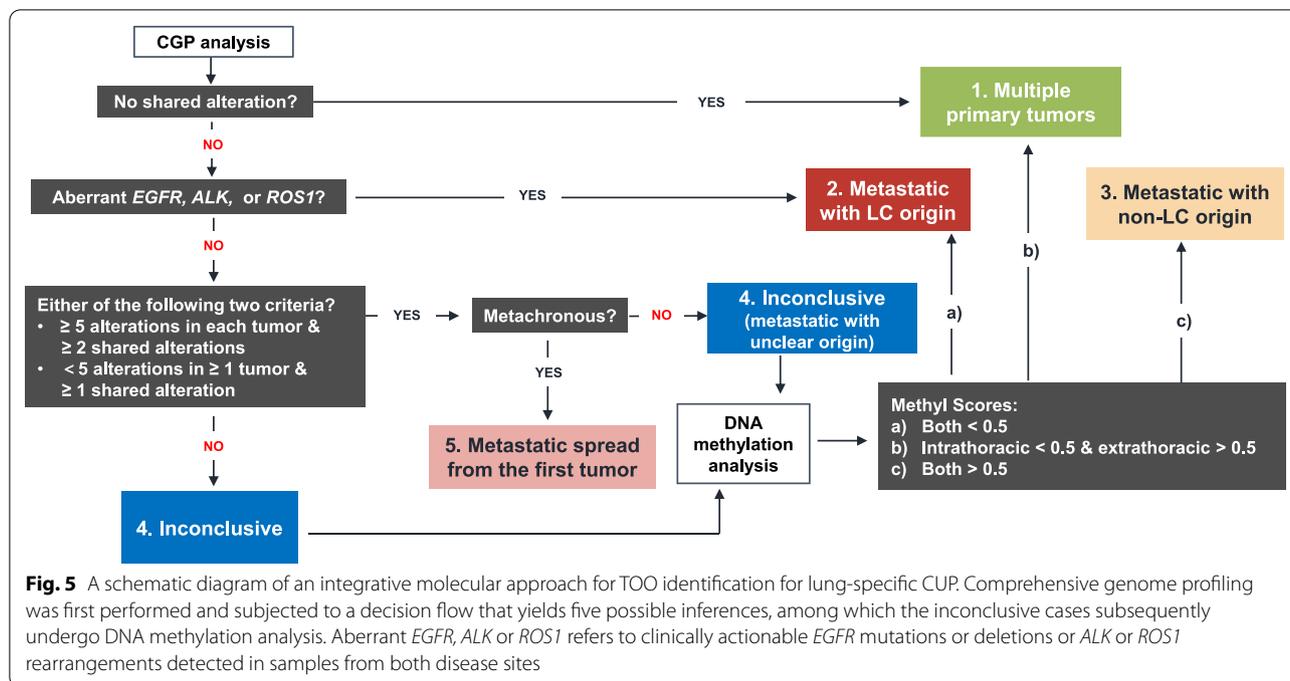
In this study, we set out to establish a multidisciplinary approach for TOO identification for lung-specific CUPs. We included 16 CUP patients and applied two molecular analyses in parallel to compare their capability of unambiguous inference and inter-method concordance (Fig. 2). The two methods achieved 100% concordance



when TOO can be clearly identified with DNA methylation analysis (Table 2), which was further supported by survival outcomes (Fig. 4). Based on these promising results, we propose an integrative molecular approach for patients with lung-specific CUP.

As illustrated in Fig. 5, it is suggested that CGP be performed first, the resulting genomic profiles subjected a decision workflow and leading to one of four possible outcomes per the illustrated decision, and that DNA methylation be conducted only when CGP was

inconclusive. This strategy was preferred over methylation profiling alone out of comprehensive consideration of relevant factors. A considerable proportion of lung-specific CUP patients carry tumors of pulmonary origin in the metastatic or the multiple primary setting, most of whom would undergo genomic profiling. Other cancer types such as colorectal and gastric cancer can also benefit from molecular testing that informs treatment biomarkers such as microsatellite stability, tumor mutation burden, and *HER2* amplification. Therefore, CGP would



be suggested for a sizable of lung-specific CUP patients once methylation analysis determines the TOO. In addition, the CGP-time lag approach used in this study gave an impressive performance of unambiguous TOO inference for nearly 70% analyzed patients, all of whom had matching inferences by methylation analysis. Taking in these considerations, we placed CGP as a first step and methylation analysis as a need-based option to maximize cost effectiveness. A possible caveat of this strategy, however, is the additional turnaround time for methylation analysis for the patients who turn out in need of it. It is therefore suggested to establish a TOO-specific sequencing program that requires greater sample amount deposited and extracted for DNA in advance, and initiates methylation profiling once CGP data are found inconclusive, thereby eliminating the need for transportation and DNA extraction of a second tumor sample. Such a program would need greater sample amount deposited in advance. In our study, 30 ng of DNA was used for methylation, which was similar to the amount for CGP and would be feasible for most cases, since a median of 2710 ng DNA (range 370–6280) can be extracted from a 20-gauge core needle biopsy [27].

CGP has been used extensively in LC to distinguish multiple primary from intrapulmonary metastasis, which hinges on elucidation of tumor clonality [13, 14]. Compared with smaller panels, large gene panels could reduce the chance of assigning clonality to tumor pairs that shared alteration by chance. Indeed, using panels of up to

468 genes, a study showed that the odds of sharing hot-spot mutations was practically nil between different primary lung tumors [13]. In this work, we applied a larger panel that targeted 520 genes to differentiate intra- and extrathoracic tumors. As expected, CGP alone was able to result in unambiguous inference of tumor clonality in all 16 (100%) patients and tumor origin for 7 (43.8%; i.e. Class I patients), including 4 with multiple primary and 3 with metastatic LC based on presence of highly specific LC drivers (actionable *EGFR* mutations: 2, *EML4-ALK* rearrangement: 1). Next, we designed the criterion of metachronous tumors based on the high propensity for metachronous tumors to result from metastatic spread [13], thereby incorporating clinical information typically available for CUP patients. This step led to TOO assignment for nearly half (4/9) of the remaining, undetermined cases. These results suggest the utility of integrating multidisciplinary evidence in tumor origin prediction.

The remarkable tissue specificity of epigenetic characteristics also been exploited for TOO prediction, which is featured by tissue origin classification based on DNA methylation patterns [28, 29]. In the EUCUP study, DNA methylation profiling predicted TOO in 87% (188/216) patients with CUP [15]. Patients treated with site-specific therapy showed significantly improved OS compared with those receiving empiric therapy. In our work, DNA methylation analysis achieved an even higher inference rate of 100%. Moreover, of the 9 patients assigned a tumor origin by both methods, CGP-time lag

and DNA methylation analyses yielded matching results for all. To our knowledge, this study is the first to evaluate the concordance level between TOO determination via genome and DNA methylation profiling and report 100% concordance. Accuracy of TOO assignment was supported by the survival outcomes. Survival analysis showed significantly better OS in patients with putative multiple primary tumors than those with metastatic cancer. Additionally, we highlighted two cases in which the clinical responses to site-specific treatment corroborated our inferences.

Despite the promising results, this study has several limitations. First, despite the promising inter-analysis concordance and significantly improved OS, the small cohort size in this retrospective study warranted further validation of these findings. Also, as patients were treated with the physician's choice of therapy, randomized studies are needed to characterize whether and how well the proposed integration could bring clinical benefits to CUP patients.

Conclusion

In this study, we evaluated the performance and level of concordance between the two molecular analyses for TOO identification in CUP. Based on the promising results, we propose an integrative strategy that combines the two methods and clinical evidence. More clinical validation and randomized trials are warranted to further characterize the value of the proposed approach in managing lung-specific CUPs.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-022-03362-2>.

Additional file 1. Comprehensive genomic profiling analysis and targeted methylation sequencing and data preprocessing.

Additional file 2: Table S1. Treatment regimens and survival outcomes for each of the 16 lung-specific CUP patients analyzed in this study.

Acknowledgements

We would like to thank the patients and their families for their firm support. We are also grateful to the staff at Burning Rock Biotech, including Xiao Zou, Songan Chen, and Haiyan Li, for technical assistance.

Authors' contribution

KYC and YF designed, wrote, and edited the manuscript; analyzed the data; and finished the figures. FRZ, XQY and ZYH revised the manuscript. LG, YJX, HL, and SZY analyzed the data. All authors read and approved the final manuscript.

Funding

This study was supported by the Basic Public Foundation of Zhejiang Province of China (No. LGF18H160017).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the Institutional Review Board of Zhejiang Cancer Hospital (No. IRB-2021-54) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent for publication

All patients had provided written informed consent for participating in the study.

Competing interests

None of the authors reported a conflict of interest related to the study.

Author details

¹The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou 310022, China. ²Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou 310022, China. ³Department of Thoracic Medical Oncology, Zhejiang Cancer Hospital, Hangzhou 310022, China. ⁴Department of Breast Surgery, Zhejiang Cancer Hospital, Hangzhou 310022, China. ⁵Department of Clinical Trial, Zhejiang Cancer Hospital, Hangzhou 310022, China. ⁶Department of Thoracic Medical Oncology, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou 310022, China.

Received: 12 September 2021 Accepted: 26 March 2022

Published online: 05 April 2022

References

- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin.* 2021;71(1):7–33.
- Riihimäki M, Hemminki A, Fallah M, Thomsen H, Sundquist K, Sundquist J, Hemminki K. Metastatic sites and survival in lung cancer. *Lung Cancer.* 2014;86(1):78–84.
- Desch CE, Benson AB 3rd, Somerfield MR, Flynn PJ, Krause C, Loprinzi CL, Minsky BD, Pfister DG, Virgo KS, Petrelli NJ. Colorectal cancer surveillance: 2005 update of an American Society of Clinical Oncology practice guideline. *J Clin Oncol.* 2005;23(33):8512–9.
- van den Broek JJ, van Gestel T, Kol SQ, van Geel AM, Geenen RWF, Schreurs WH. Dealing with indeterminate pulmonary nodules in colorectal cancer patients; a systematic review. *Eur J Surg Oncol.* 2021;47(11):2749–56.
- Choi JY, Ha TK, Kwon SJ. Clinicopathologic characteristics of gastric cancer patients according to the timing of the recurrence after curative surgery. *J Gastric Cancer.* 2011;11(1):46–54.
- Ki EY, Lee KH, Park JS, Hur SY. A clinicopathological review of pulmonary metastasis from uterine cervical cancer. *Cancer Res Treat.* 2016;48(1):266–72.
- Lee MS, Sanoff HK. Cancer of unknown primary. *BMJ.* 2020;371:m4050.
- Larson MH, Pan W, Kim HJ, Mauntz RE, Stuart SM, Pimentel M, Zhou Y, Knudsgaard P, Demas V, Aravanis AM, Jamshidi A. A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. *Nat Commun.* 2021;12(1):2357.
- Danilova L, Wrangle J, Herman JG, Cope L. DNA-methylation for the detection and distinction of 19 human malignancies. *Epigenetics.* 2021;17(2):191–201.
- Kato S, Alsafar A, Walavalkar V, Hainsworth J, Kurzrock R. Cancer of unknown primary in the molecular era. *Trends Cancer.* 2021;7(5):465–77.
- Hayashi H, Kurata T, Takiguchi Y, Arai M, Takeda K, Akiyoshi K, Matsumoto K, Onoe T, Mukai H, Matsubara N, Minami H, Toyoda M, Onozawa Y, Ono A, Fujita Y, Sakai K, Koh Y, Takeuchi A, Ohashi Y, Nishio K, Nakagawa K. Randomized Phase II trial comparing site-specific treatment based on

- gene expression profiling with carboplatin and paclitaxel for patients with cancer of unknown primary site. *J Clin Oncol.* 2019;37(7):570–9.
12. Hayashi H, Takiguchi Y, Minami H, Akiyoshi K, Segawa Y, Ueda H, Iwamoto Y, Kondoh C, Matsumoto K, Takahashi S, Yasui H, Sawa T, Onozawa Y, Chiba Y, Togashi Y, Fujita Y, Sakai K, Tomida S, Nishio K, Nakagawa K. Site-specific and targeted therapy based on molecular profiling by next-generation sequencing for cancer of unknown primary site: a nonrandomized phase 2 clinical trial. *JAMA Oncol.* 2020;6(12):1931–8.
 13. Chang JC, Alex D, Bott M, Tan KS, Seshan V, Golden A, Sauter JL, Buonocore DJ, Vanderbilt CM, Gupta S, Desmeules P, Bodd FM, Riely GJ, Rusch VW, Jones DR, Arcila ME, Travis WD, Ladanyi M, Rekhtman N. Comprehensive next-generation sequencing unambiguously distinguishes separate primary lung carcinomas from intrapulmonary metastases: comparison with standard histopathologic approach. *Clin Cancer Res.* 2019;25(23):7113–25.
 14. Wang X, Gong Y, Yao J, Chen Y, Li Y, Zeng Z, Lu Y, Song L. Establishment of criteria for molecular differential diagnosis of MPLC and IPM. *Front Oncol.* 2020;10:614430.
 15. Moran S, Martínez-Cardús A, Sayols S, Musulén E, Balañá C, Estival-Gonzalez A, Moutinho C, Heyn H, Diaz-Lagares A, de Moura MC, Stella GM, Comoglio PM, Ruiz-Miró M, Matias-Guiu X, Pazo-Cid R, Antón A, Lopez-Lopez R, Soler G, Longo F, Guerra I, Fernandez S, Assenov Y, Plass C, Morales R, Carles J, Bowtell D, Mileskin L, Sia D, Tothill R, Taberner J, Llovet JM, Esteller M. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol.* 2016;17(10):1386–95.
 16. Liu H, Chen J, Chen H, Xia J, Wang O, Xie J, Li M, Guo Z, Chen G, Yan H. Identification of the origin of brain metastases based on the relative methylation orderings of CpG sites. *Epigenetics.* 2021;16(8):908–16.
 17. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol.* 2020;31(6):745–59.
 18. Koch A, Joosten SC, Feng Z, de Ruijter TC, Draht MX, Melotte V, Smits KM, Veeck J, Herman JG, Van Neste L, Van Criekinge W, De Meyer T, van Engeland M. Analysis of DNA methylation in cancer: location revisited. *Nat Rev Clin Oncol.* 2018;15(7):459–66.
 19. Liang N, Li B, Jia Z, Wang C, Wu P, Zheng T, Wang Y, Qiu F, Wu Y, Su J, Xu J, Xu F, Chu H, Fang S, Yang X, Wu C, Cao Z, Cao L, Bing Z, Liu H, Li L, Huang C, Qin Y, Cui Y, Han-Zhang H, Xiang J, Liu H, Guo X, Li S, Zhao H, Zhang Z. Ultrasensitive detection of circulating tumour DNA via deep methylation sequencing aided by machine learning. *Nat Biomed Eng.* 2021;5(6):586–99.
 20. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, Nicholson AG, Groome P, Mitchell A, Bolejack V. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (Eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol.* 2016;11(1):39–51.
 21. Xu Y, Li H, Huang Z, Chen K, Yu X, Sheng J, Zhang HH, Fan Y. Predictive values of genomic variation, tumor mutational burden, and PD-L1 expression in advanced lung squamous cell carcinoma treated with immunotherapy. *Transl Lung Cancer Res.* 2020;9(6):2367–79.
 22. Xu T, Zhang Y, Zhang J, Qi C, Liu D, Wang Z, Li Y, Ji C, Li J, Lin X, Hou T, Liu H, Zhang L, Han-Zhang H, Shen L, Wang X. Germline profiling and molecular characterization of early onset metastatic colorectal cancer. *Front Oncol.* 2020;10:568911.
 23. Duan J, Ge M, Peng J, Zhang Y, Yang L, Wang T, Qin T, Yuan R, Zhang Y, Cheng W. Application of large-scale targeted sequencing to distinguish multiple lung primary tumors from intrapulmonary metastases. *Sci Rep.* 2020;10(1):18840.
 24. Sui J, Wu X, Wang C, Wang G, Li C, Zhao J, Zhang Y, Xiang J, Xu Y, Nian W, Cao F, Yu G, Lou Z, Hao L, Liu L, Li B, Zhang Z, Cai S, Liu H, Lan P, Zhang W. Discovery and validation of methylation signatures in blood-based circulating tumor cell-free DNA in early detection of colorectal carcinoma: a case-control study. *Clin Epigenetics.* 2021;13(1):26.
 25. Tang W, Wan S, Yang Z, Teschendorff AE, Zou Q. Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics.* 2018;34(3):398–406.
 26. Scagliotti GV, Parikh P, von Pawel J, Biesma B, Vansteenkiste J, Manegold C, Serwatowski P, Gatzemeier U, Digumarti R, Zukin M, Lee JS, Mellemaard A, Park K, Patil S, Rolski J, Goksel T, de Marinis F, Simms L, Sugarman KP, Gandara D. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol.* 2008;26(21):3543–51.
 27. Silk MT, Mikkilineni N, Silk TC, Zabor EC, Ostrovnyaya I, Hakimi AA, Hsieh JJ, Ziv E, Rekhtman N, Solomon SB, Durack JC. Prospective evaluation of unprocessed core needle biopsy DNA and RNA yield from lung, liver, and kidney tumors: implications for cancer genomics. *Anal Cell Pathol.* 2018;2018:2898962.
 28. Modhukur V, Sharma S, Mondal M, Lawarde A, Kask K, Sharma R, Salumets A. Machine learning approaches to classify primary and metastatic cancers using tissue of origin-based DNA methylation profiles. *Cancers.* 2021;13(15):3768.
 29. Abraham J, Heimberger AB, Marshall J, Heath E, Drabick J, Helmstetter A, Xiu J, Magee D, Stafford P, Nabhan C, Antani S, Johnston C, Oberley M, Korn WM, Spetzler D. Machine learning analysis using 77,044 genomic and transcriptomic profiles to accurately predict tumor type. *Transl Oncol.* 2021;14(3):101016.

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

