Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis

Sebastian Moran, Anna Martínez-Cardús, Sergi Sayols, Eva Musulén, Carme Balañá, Anna Estival-Gonzalez, Cátia Moutinho, Holger Heyn, Angel Diaz-Lagares, Manuel Castro de Moura, Giulia M Stella, Paolo M Comoglio, Maria Ruiz-Miró, Xavier Matias-Guiu, Roberto Pazo-Cid, Antonio Antón, Rafael Lopez-Lopez, Gemma Soler, Federico Longo, Isabel Guerra, Sara Fernandez, Yassen Assenov, Christoph Plass, Rafael Morales, Joan Carles, David Bowtell, Linda Mileshkin, Daniela Sia, Richard Tothill, Josep Tabernero, Josep M Llovet, Manel Esteller

Summary

Background **Cancer of unknown primary ranks in the top ten cancer presentations and has an extremely poor prognosis. Identification of the primary tumour and development of a tailored site-specific therapy could improve the survival of these patients. We examined the feasability of using DNA methylation profiles to determine the occult original cancer in cases of cancer of unknown primary.**

Methods **We established a classifier of cancer type based on the microarray DNA methylation signatures (EPICUP) in a training set of 2790 tumour samples of known origin representing 38 tumour types and including 85 metastases. To validate the classifier, we used an independent set of 7691 known tumour samples from the same tumour types that included 534 metastases. We applied the developed diagnostic test to predict the tumour type of 216 well-characterised cases of cancer of unknown primary. We validated the accuracy of the predictions from the EPICUP assay using autopsy examination, follow-up for subsequent clinical detection of the primary sites months after the initial presentation, light microscopy, and comprehensive immunohistochemistry profiling.**

Findings **The tumour type classifier based on the DNA methylation profiles showed a 99·6% specificity (95% CI 99·5–99·7), 97·7% sensitivity (96·1–99·2), 88·6% positive predictive value (85·8–91·3), and 99·9% negative predictive value (99·9–100·0) in the validation set of 7691 tumours. DNA methylation profiling predicted a primary cancer of origin in 188 (87%) of 216 patients with cancer with unknown primary. Patients with EPICUP diagnoses who received a tumour type-specific therapy showed improved overall survival compared with that in patients who received empiric therapy (hazard ratio [HR] 3·24, p=0·0051 [95% CI 1·42–7·38]; log-rank p=0·0029).**

Interpretation **We show that the development of a DNA methylation based assay can significantly improve diagnoses of cancer of unknown primary and guide more precise therapies associated with better outcomes. Epigenetic profiling could be a useful approach to unmask the original primary tumour site of cancer of unknown primary cases and a step towards the improvement of the clinical management of these patients.**

Funding **European Research Council (ERC), Cellex Foundation, the Institute of Health Carlos III (ISCIII), Cancer Australia, Victorian Cancer Agency, Samuel Waxman Cancer Research Foundation, the Health and Science Departments of the Generalitat de Catalunya, and Ferrer.**

Introduction

Cancer of unknown primary accounts for about 3–9% of all cancer diagnoses,[1] and in the USA alone, more than 80 000 patients receive a diagnosis of cancer of unknown primary every year.[2] With a median age at presentation of 60 years, cancers of unknown primary are the fourth most common cause of cancer-related deaths worldwide.[3] Cancers of unknown primary are a molecularly heterogeneous group of cancers, in which there has not yet been elucidation of the biological mechanisms that allow the primary site to remain obscure after metastasis. No common molecular signature has been identified that produces this particular clinical phenotype, and cancers of unknown primary present a wide variety of mutations and genomic alterations.[4] From a clinical standpoint, the prognosis for patients with cancer of unknown primary is poor: patients attain a median survival of 9 months (95% CI 8·3–10·0)[5] after diagnosis and only 25% survive for 1 year or more.[6] For most patients with cancer of unknown primary, recommended treatments involve empirical chemotherapy—defined as the chemotherapy that the oncologist think will work best based on their experience treating other people with similar characteristics—usually with a taxane plus platinum, or gemcitabine plus a platinum regimen,[7–9] which produce the described modest clinical benefit. However, accurate identification of the primary tumour type and subsequent treatment with site-specific therapy could result in improved survival.[10–13]

If the initial assessment of a patient with a cancer of unknown primary, which usually involves CT scanning and specific signs or symptoms, is uninformative, the first attempt to identify a tissue of origin relies on pathological assessment, including an immunohistochemical examination. Several immunohistochemical panels have been developed for the diagnoses of cancer of unknown primary, but even after the full diagnostic

Lancet Oncol 2016

Published Online
August 26, 2016
http://dx.doi.org/10.1016/S1470-2045(16)30297-2

See Online/Comment
http://dx.doi.org/10.1016/S1470-2045(16)30377-1

Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet, Barcelona, Catalonia, Spain (S Moran MSc, A Martínez-Cardús PhD, S Sayols MSc, C Moutinho PhD, H Heyn PhD, A Diaz-Lagares PhD, M Castro de Moura MSc, Prof M Esteller MD); **Department of Pathology, Hospital Universitari Germans Trias i Pujol, C/ Ctra de Canyet s/n, Badalona, Barcelona, Catalonia, Spain** (E Musulén MD); **Medical Oncology, Catalan Institute of Oncology (ICO), University Hospital Germans Trias i Pujol, Badalona, Barcelona, Catalonia, Spain** (C Balañá MD, A Estival-Gonzalez MD); **Cardiothoracic and Vascular Department, Pneumology Unit, IRCCS Policlinico San Matteo Foundation, Pavia, Italy** (G M Stella MD); **Institute for Cancer Research at Candiolo, Candiolo, Italy** (Prof P M Comoglio MD); **IRBLleida Biobank, Lleida, Catalonia, Spain** (M Ruiz-Miró PhD); **Department of Pathology and Molecular Genetics/Oncologic Pathology Group, Hospital Universitari Arnau de Vilanova, Universitat de Lleida, IRBLleida, Lleida, Catalonia, Spain** (X Matias-Guiu MD); **Medical Oncology Service, Hospital Miguel Servet, Zaragoza, Spain** (R Pazo-Cid MD, A Antón MD); **Medical Oncology Service, Complejo Hospitalario Universitario de Santiago, Santiago de Compostela, Spain** (R Lopez-Lopez MD); **Medical Oncology, Catalan Institute of**

Oncology (ICO), Hospital Duran i Reynals, L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain (G Soler MD); Medical Oncology Service, Hospital Universitario Ramon y Cajal, Madrid, Spain (F Longo MD); Biobanco Vasco, Hospital Universitario de Araba, Vitoria, Spain (I Guerra MD); Biobanco Vasco, Hospital Universitario de Basurto, Bilbao, Spain (S Fernandez MD); Division of Epigenomics and Cancer Risk Factors at the German Cancer Research Center (DKFZ), Heidelberg, Germany (Y Assenov PhD, Prof C Plass PhD); Oncology Department, Vall d'Hebron University Hospital, Universitat Autònoma de Barcelona (UAB), Barcelona, Catalonia, Spain (R Morales MD, J Carles MD, J Tabernero MD); Oncology Department, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Catalonia, Spain (R Morales, J Carles, J Tabernero); Liver Cancer Translational Research Laboratory, Barcelona Clinic Liver Cancer (BCLC) Group, Liver Unit, IDIBAPS, Hospital Clínic, CIBERehd, Barcelona, Catalonia, Spain (Prof J M Llovet MD); School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain (Prof J M Llovet, Prof M Esteller); Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain (Prof J M Llovet, Prof M Esteller); The Peter MacCallum Cancer Centre, Melbourne, VIC, Australia (Prof D Bowtell PhD, L Mileshkin MD, R Tothill PhD); The Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria, Australia (Prof D Bowtell, L Mileshkin); The Department of Pathology, University of Melbourne, Parkville, VIC, Australia (Prof D Bowtell, L Mileshkin, R Tothill); and Liver Cancer Program, Division of Liver Diseases, Tisch Cancer Institute, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA (D Sia PhD, Prof J M Llovet)

Correspondence to:
Prof Manel Esteller, Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), 08908 L'Hospitalet, Barcelona, Catalonia, Spain
mesteller@idibell.cat

Research in context

Evidence before this study

This study was initiated based on our preclinical data showing that DNA methylation patterns are tumour type specific, a finding that could be helpful in the identification of the site of origin of cancer of unknown primary. Additionaly, we searched PubMed on April 8, 2016, unrestricted by language or date limits, to identify scientific literature focused on the diagnosis and therapies of cancer of unknown primary using the search term "cancer of unknown primary". We also searched abstracts from the American Society of Clinical Oncology and the European Society for Medical Oncology. We found no studies that examined the use of epigenetic profiling to improve the clinical management of cancer of unknown primary.

Added value of this study

Our findings show that a classifier of cancer type based on microarray DNA methylation signatures shows a high specificity, sensitivity, positive predictive value, and negative predictive value for the prediction of the original primary tumour site. We validated the accuracy of the test using autopsy examination, subsequent clinical detection of the primary site, light microscopy, and comprehensive immunohistochemistry profiling. Additionally, our results suggest that patients with cancer of unknown primary who received a tumour type-specific therapy showed improved overall survival compared with that in patients who received empiric therapy. The test also suggested the presence of actionable targets such as *HER2* and *C-MET* amplification and *EGFR* mutation.

Implications of all the available evidence

The results of this study could change the diagnosis of patients with cancer of unknown primary where the approaches routinely used to determine the tissue of origin provide conclusive results in only 25% of cases. Our data support the use of epigenetic profiling to significantly improve cancer of unknown primary diagnoses and guide more precise therapies associated with better outcomes.

work-up, the primary site of a cancer of unknown primary remains unknown in about 75% of patients.[1] A post-mortem examination is done in only a few patients with cancer of unknown primary, in which a complete autopsy only reveals 55–85% of the primaries.[6] Our increasing understanding of cancer biology has prompted the search for molecular markers that, being present in the cancer of unknown primary, might retain the signature of the putative primary origin. In this regard, the use of expression microarray-based classifiers has achieved a prediction accuracy of the primary site in about 75% of patients.[1,14,15] However, the limitations in the numbers, types, and subtypes of tumours included in these assays, in addition to the required amount and state of preservation of the studied biological material, and the cost of the procedure, warrant further development of complementary diagnostic instruments for cancer of unknown primary. In this regard, we examined DNA methylation,[16–18] a stable marker of DNA that has already been clinically successful in the pharmacogenetic management of gliomas[19,20] and has different profiles among distinct tumour types.[21] Herein, we attempt to diagnose the primary tumour site for all cancers of unknown primary, and we have devised a new strategy based on the DNA methylation profiles of the metastasis sample.

Methods

Patients and samples

Between March 2, 2011, and Dec 2, 2015, samples for the training (n=692) and validation (n=1948) sets were obtained from the Cancer Epigenetics and Biology Program (PEBC) of the Bellvitge Biomedical Research Institute (IDIBELL; Barcelona, Catalonia, Spain). The histopathology findings and the clinical data from the PEBC samples were obtained from the authors' institutions, according to the protocol approved by the Bellvitge University Hospital Clinical Investigation Ethics Committee (PR133/14). DNA methylation microarray data from additional tumour samples of known origin from The Cancer Genome Atlas (TCGA; (National Cancer Institute and National Human Genome Research Institute, Bethesda, MD, USA), corresponding to the tumour types studied here, were also included in the training (n=2098) set and validation (n=5743) set. Cancers of unknown primary was defined following the European Society of Medical Oncology guidelines as metastatic tumours for which the standardised diagnostic work-up failed to identify the site of origin at the time of diagnosis.[22] Paraffin-embedded tumour tissue samples from 216 patients with cancer of unknown primary were retrospectively and prospectively collected from 11 health centres from the USA, Spain, Germany, Italy, and Australia (appendix p 4). Each health centre had their own cancer of unknown primary institutional diagnostic work-up. Molecular screening of alterations in the main oncodrivers, and immunohistochemical stainings routinely analysed in clinical care were done at each participating centre, and clinical data associated with disease outcome were collected when available.

The study protocol was approved by the appropriate Ethics Committees (PR133/14). Patients gave their signed, informed consent when required, and applicable according to the institutional review board at each institute.

Histopathological evaluation and molecular analysis

Histology-guided tumour-type classification of cancer of unknown primary involved review by a pathologist of the tumour's morphological appearance under light

microscopy, as well as immunohistochemical findings, including cytokeratin 7 (CK7), cytokeratin 20 (CK20), vimentin, epithelial membrane antigens, and S-100 expression. Further detailed immunohistochemical classification was done as described in the appendix (p 3). According to the predicted cancer of unknown primary tumour type, we assessed the possible presence of *HER2* or *C-MET* gene amplification, *ALK* and *ROS1* translocations, and oncogenic point mutations in *EGFR*.

DNA methylation microarray, data analysis, and algorithm development

DNA from fresh-frozen samples was extracted with the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), while four 10 µm sections of formalin-fixed, paraffin-embedded (FFPE) blocks were processed using the EZNA FFPE DNA kit (Omega Bio-tek, Norcross, GA, USA). For the DNA methylation microarray study, 300 ng of FFPE DNA, or 600 ng of FF DNA were randomly distributed on a 96-well plate, and processed with the EZ-96 DNA Methylation kit (Zymo Research Corp, Irvine, CA, USA). Bisulfite-converted DNA (bs-DNA) from FFPE samples was processed as detailed in the Infinium FFPE Restoration guide (Illumina, San Diego, CA, USA).[23] Microarray hybridisation and scanning were done as previously described.[24] Raw data (intensity data, intensity data files) were normalised one at a time by normalising each sample against a normalising set, consisting of a previously defined subset of 100 of the training samples chosen by PEBC. A three-step-based normalisation procedure was done using the lumi package available for bioconductor, within the R statistical environment, consisting of colour bias adjustment, background level adjustment, and quantile normalisation across arrays. The methylation levels (β values) for each of the 485 577 CpG sites were calculated as the ratio of methylated signal to the sum of methylated and unmethylated signals plus 100. After the normalisation step, probes related to X and Y chromosomes were removed, as were those probes whose 10 bases nearest the interrogated site contained a SNP, as noted in the product description file.

For each of the probes resulting from the normalisation, an analysis of variance was done after categorisation of each of the samples of the training set into one of the 38 tumour types. Resulting p values were corrected using the Bonferroni method, and Tukey's honest significant difference post-hoc test was applied. CpGs that were specific for at least one tumour type were selected ($\Delta\beta>0\cdot2$, $p<0\cdot01$). The importance of the variables for each of the resulting CpG sites was estimated with the mean decrease in accuracy with a random forest machine learning method available in the R environment (n-tree=1000). The resulting CpG sites were ranked according to their informativeness in separating the groups of tumoural types. Nested models (n-tree=1000) were constructed using the variables ranked from the most to the least important, until the predictive power of the model stagnated. Finally, CpGs whose information was redundant were excluded, by adding important CpGs, one by one, and excluding those that did not add predictive value to the model. A random forest classifying algorithm was created using the obtained CpGs.

Normalised samples of the validation set were used to assess the robustness of the classifying algorithm. Samples were passed blind by the classifier algorithm and the results compared with the initial classification of the sample into one of the 38 tumour types, corresponding to the most common human cancers. Out of the 38 similarity scores reported by the algorithm for each sample, values above a threshold (similarity score ≥0·12) were considered positive results. If the higher positive result agreed with the reported diagnosis of the sample a true positive result was reported, whereas if values above the similarity threshold did not match the reported diagnosis, they were considered true negatives. False positive results (positive algorithm results that did not match the reported diagnosis) and false negative results (reported diagnosis among the negative algorithm results) were used accordingly. A confusion matrix was generated for each tumoural type. 95% CIs for proportions were calculated according to the efficient-score method. The geometrical mean of the tumour type statistics was computed from the overall measurements. Further data analysis and algorithm development are described in the appendix (p 2).

Statistical analysis

Survival analysis examined the association between disease outcome of cancer of unknown primary and the type of chemotherapy. The associations between categorical variables were analysed by χ^2 tests or Fisher's exact test, as appropriate. Kaplan-Meier plots and the log-rank test were used to estimate the effect of the administration of a specific treatment in progression-free survival and overall survival. The associations of clinical parameters with overall survival (time from diagnosis to death) were assessed with univariate and multivariable Cox proportional hazards regression models. All statistical tests were two-sided and values of $p<0\cdot05$ were considered statistically significant. Data were analysed with IBM SPSS (version 20) software.

Role of the funding source

The funders had no role in the study's conduct, design, data collection and analysis, data interpretation, or writing the report. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Results

To obtain the basal DNA methylation landscapes associated with 38 tumour types, we analysed 10 481 tumour samples of known origin: 2790 in the

For **The Cancer Genome Atlas** see http://cancergenome.nih.gov/

See **Online** for appendix

training set and 7691 in the validation set (table 1). Each tumour type had at least 25 cases adding training plus validation set. We studied 692 tumour samples of known origin from the PEBC cohort to establish a reference dataset of DNA methylation profiles associated with 38 tumour types that included those typically associated with cancer of unknown primary (table 1). We obtained the DNA methylation signature for all these cases by using a comprehensive microarray that interrogated the methylation status of around half a million CpG sites in the human genome.[24] The same epigenetic platform has been used in TCGA to analyse the DNA methylomes of a large number of tumour types, so we were able to add this publicly available information to that from our training cohort. The TCGA cases included 2098 tumour samples of known origin (table 1) of the same 38 tumour types as in our cohort. Thus, the tumour type classifier was trained using a total of 2790 tumour samples of known origin.

	Training sample			Validation sample		
	PEBC	TCGA	Total	PEBC	TCGA	Total
Acute lymphoblastic leukaemia	50	0	50	30	0	30
Acute myeloid leukaemia	50	50	100	18	144	162
Adrenocortical carcinoma	0	50	50	0	30	30
Bladder urothelial carcinoma	20	80	100	11	179	190
Brain lower-grade glioma	20	80	100	28	482	510
Breast carcinoma	12	88	100	97	660	757
Cervical squamous carcinoma	0	100	100	5	111	116
Chronic lymphocytic leukaemia	32	0	32	10	0	10
Colon carcinoma	30	70	100	618	222	840
Cutaneous lymphoma	15	0	15	12	0	12
Endometrial carcinoma	20	80	100	30	359	389
Oesophageal carcinoma	14	86	100	2	40	42
Head and neck squamous cell carcinoma	13	87	100	5	430	435
Hepatocellular carcinoma	50	50	100	179	233	412
Lymphoid neoplasm diffuse large B-cell lymphoma	0	18	18	9	10	19
Meningioma	15	0	15	11	0	11
Mesothelioma	0	50	50	0	37	37
Multiple myeloma	50	0	50	24	0	24
Neuroendocrine carcinoma	50	0	50	18	0	18
Non-small-cell lung carcinoma	40	60	100	498	765	1263
Non-seminomatous germ cell tumors	2	48	50	10	29	39
Ovarian carcinoma	11	7	18	25	3	28
Pancreatic carcinoma	12	88	100	15	42	57
Pheochromocytoma	0	100	100	0	84	84
Prostate carcinoma	15	85	100	29	344	373
Rectal adenocarcinoma	0	75	75	0	21	21
Renal tumour chromophobe	10	40	50	5	26	31
Renal tumour clear cell	8	92	100	12	209	221
Renal tumour papillary	10	90	100	13	92	105
Retinoblastoma	20	0	20	10	0	10
Sarcoma	31	69	100	35	156	191
Seminoma	3	47	50	15	27	42
Skin cutaneous melanoma	20	80	100	98	296	394
Small-cell lung carcinoma	46	1	47	15	1	16
Stomach carcinoma	10	90	100	37	236	273
Thymoma	0	100	100	0	24	24
Thyroid carcinoma	5	95	100	17	413	430
Uveal melanoma	8	42	50	7	38	45
Total	692	2098	2790	1948	5743	7691

PEBC=Cancer Epigenetics and Biology Program (Barcelona, Catalonia, Spain). TCGA=The Cancer Genome Atlas.

Table 1: **Training (n=2790) and validation (n=7691) sample distributions by tumour type**

To validate the classifier, we used an independent set of 1948 tumour samples of known origin (table 1) from the PEBC cohort that had the same distribution in the same 38 tumour types as in the training set. We obtained the DNA methylation profile of each case using the same method, genomic platform and analytical procedure as for the training cohort. We also consulted the TCGA databases and incorporated in our study 5743 tumour samples of known origin representing the tumour types included in the training set (table 1). Thus, our validation set to predict tumour type consisted of 7691 known tumour samples (table 1).

The tumour type classifier based on the DNA methylation profiles, hereafter referred to as EPICUP, showed a 99·6% (95% CI 99·5–99·7) specificity (true negative rate) and 97·7% (96·1–99·2) sensitivity (true positive rate) for determining the tumour type of the studied 7691 samples. EPICUP also showed a 88·6% (95% CI 85·8–91·3) positive predictive value and 99·9% (99·9–100) negative predictive value in the 7691 samples in the validation set.

We used three types of experiments to evaluate assay reproducibility and the reliability of the DNA methylation classifier. First, 11 samples (randomly selected with PEBC) of three known tumour types (colorectal carcinoma, breast carcinoma, and cutaneous melanoma) were hybridised to the DNA methylation microarray in three batches at different times, giving rise to the same predicted tumour type in all cases (appendix p 11). Second, we examined in the validation cohort whether the type of DNA material, extracted from fresh frozen tissue (n=7146) or FFPE sections (n=545), affected the EPICUP classifier (appendix p 12). In all cases, the method of preservation of the DNA had no effect on the tumour type predicted by the DNA methylation classifier, confirming the reliability of the described platform for the study of archive material.[23] Finally, we examined whether the metastasis from a particular tumour type had a radically different DNA methylation profile in relation to its primary site of origin that could cause the sample to be misclassified. This is particularly pertinent because it reflects the clinical circumstances of cancer of unknown primary. The validation cohort included 534 metastases where the tumour type of the primary site was known, representing 21 different origins (appendix p 13). We found that EPICUP predicted the correct tumour type for 501 (94%) of 534 metastases.

Table 2 shows the characteristics of the 216 patients with cancer of unknown primary in whom the diagnostic test was applied. These cases of cancer of unknown primary were diagnosed at the participating centres between Jan 1, 1998, and Oct 28, 2015. The cohort shared the usual clinicopathological features observed in reported cases of cancer of unknown primary:[12,15] a median age of 63 years (range 29–89); a similar distribution in male and female cases; and a predominance of adenocarcinomas and carcinomas (table 2). For the 114 cases of cancer of unknown primary for which clinical data associated with disease outcome were available (appendix p 5), we observed a median overall survival of 8·1 months (95% CI 0·1–143·4). These patients were not given site-specific therapy based on the results of the EPICUP assay. The immunohistochemical evaluation received by the 216 patients with cancer of unknown primary is shown in the appendix (pp 6–10).

	Patients with cancer of unknown primary site (n=216)
Sex	
Male	120 (56%)
Female	96 (44%)
Age, years	63 (29–89)
Diagnostic method	
Biopsy	109 (51%)
Surgery	30 (14%)
Imaging	23 (11%)
Not specified	54 (25%)
Biopsy site	
Lymph nodes	63 (29%)
Liver	44 (20%)
Bone	19 (9%)
CNS	16 (7%)
Peritoneum	11 (5%)
Skin	10 (5%)
Soft tissues	8 (4%)
Abdomen	6 (3%)
Lung	6 (3%)
Pleura	5 (2%)
Intestine	4 (2%)
Thorax	4 (2%)
Ovary	3 (1%)
Breast	2 (1%)
Other	8 (4%)
Not specified	7 (3%)
Histological diagnosis	
Adenocarcinoma/carcinoma	143 (66%)
Squamous carcinoma	39 (18%)
Undifferentiated neoplasia	26 (12%)
Sarcomatoid	1 (1%)
Not specified	7 (3%)
Metastasis sites at diagnosis	
Multiple	130 (60%)
Single	49 (23%)
Not specified	37 (17%)

Data are n (%) or median (range).

Table 2: **Clinical characteristics of patients with cancer of unknown primary site included in the study**

The DNA methylation profiling assay predicted the tissue of origin in 188 (87%) of 216 patients with cancer of unknown primary. 23 types of tissue of origin were predicted (appendix p 14). The six most commonly predicted tissues of origin were non-small-cell lung carcinoma (NSCLC; 39 [21%] of 188), head and neck squamous cell carcinoma (18 [10%]), breast carcinoma (17 [9%]), colon carcinoma (16 [9%]), hepatocellular carcinoma (14 [7%]), and pancreatic carcinoma (14 [7%]). Overall, these sites accounted for 63% of all patients. The epigenetic profile strategy can be translated to newly developed microarrays that have an expanded number of methylated sites interrogated throughout the genome,[25] but maintain those used in the EPICUP development. We found that for eight studied cases of cancer of unknown primary, EPICUP concluded the same tumour type with both epigenomic platforms (450K and 850K EPIC microarray platforms; appendix p 15), opening an avenue for further technological development.

Verification of the results of the EPICUP assay can be done in several ways. One option is to identify the primary site in an autopsy, but this is rarely done in current clinical practice. We obtained an autopsy confirmation of our EPICUP diagnosis for one case in our cancer of unknown primary cohort (table 3). The cancer of unknown primary was first diagnosed as an undifferentiated neoplasm by a biopsy of lymph node of the supraclavicular area. The DNA methylation profile established that it corresponded to a sarcoma. Notably, the autopsy found two additional metastases at the meninges and humerus that were pathologically diagnosed as sarcoma following detailed immunohistochemistry. This prompted us to reexamine the original cancer of unknown primary case using

	Samples tested (n)	Comparison with EPICUP prediction			Accuracy (%)
		Compatible	Non-compatible	Non-informative	
Necropsy	1	1	0	..	100%
Further appearance of primary tumour	38	33	5	..	87%
Light microscopy evaluation	181	174	7	..	96%
IHC with tissue-specific markers	43	31	0	12	100%

Accuracy is calculated by comparing compatible and non-compatible cases. Non-informative cases are not considered in the accuracy calculation. EPICUP=microarray DNA methylation signatures. IHC=immunohistochemistry.

Table 3: **EPICUP prediction accuracy compared with other clinical diagnostic tests**

immunohistochemistry for vimentin, because it is typically expressed by sarcomas. This led to a compatible diagnosis of sarcoma once again.

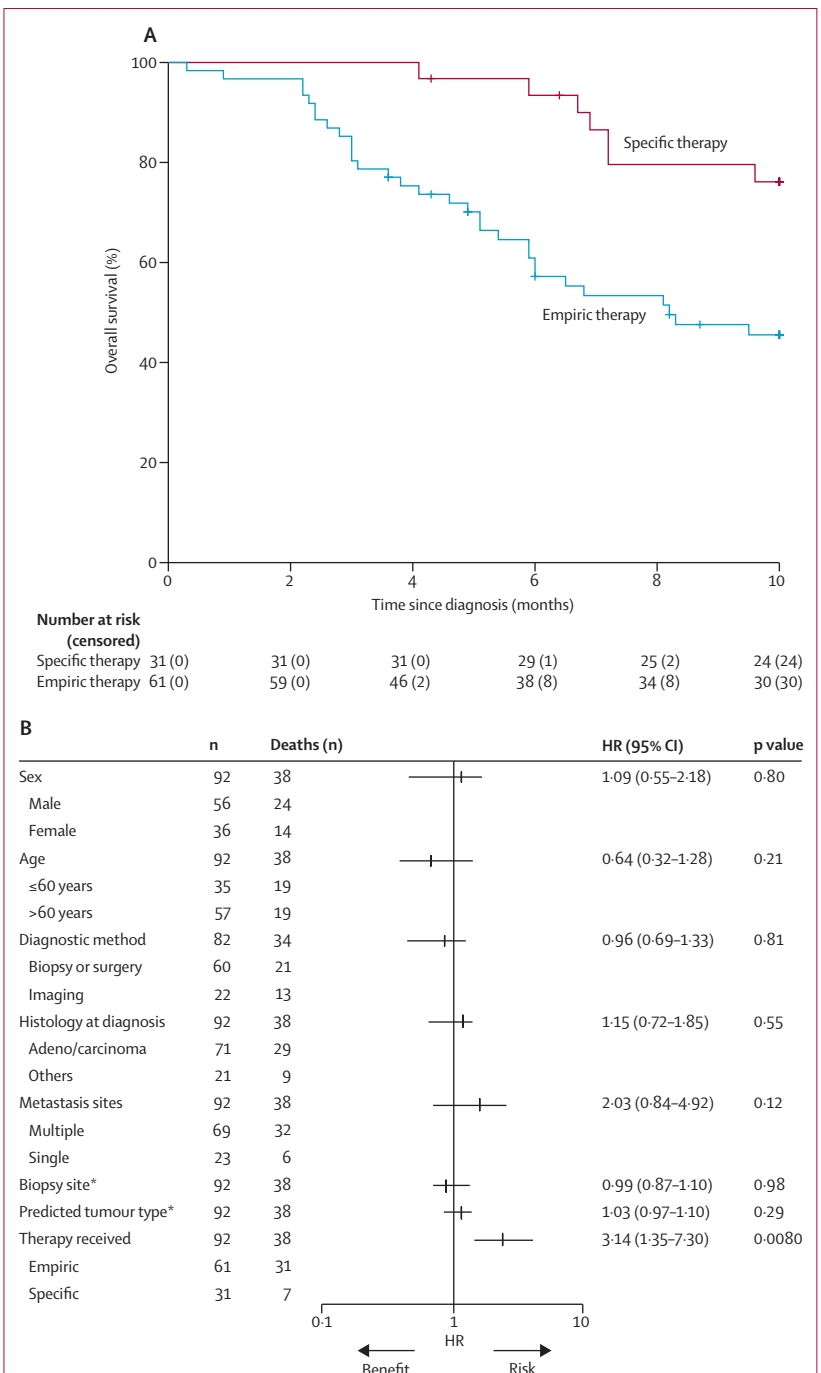

Figure: **Outcome of patients with cancer of unknown primary who receive a site-specific treatment that matches the EPICUP prediction**

(A) Kaplan-Meier curve analysis of overall survival comparing patients who received site-specific treatment according to tumour type prediction by epigenetic profiling versus empiric therapy. (B) Forest plot of the multivariable Cox regression for overall survival. Parameters with associated values of p<0·05 were considered to be independent prognostic factors. EPICUP=microarray DNA methylation signatures. HR=hazard ratio. *These comparisons were assessed in more than two groups and then only the overall result is shown.

Another scenario that offers a direct test to assess the accuracy of EPICUP prediction is the evaluation of patients with cancer of unknown primary who subsequently develop clinically detectable primary sites months after the initial presentation. We found that among the cases in which a primary tumor was found later in life, EPICUP predicted the same cancer type in 33 (87%) of 38 cases (table 3). The most commonly identified cases of cancer of unknown primary found in this manner were derived from the colon (n=11), pancreas (n=5), and breast (n=5; appendix p 16). An illustrative example includes a case of cancer of unknown primary that debuted with several affected lymph nodes, complementary negative imaging tests and uninformative immunohistochemistry results, and was treated with empirical chemotherapy. 28 months later, CT images showed thickening of the head of the pancreas and the biopsy sample provided a pancreatic cancer diagnosis—the same as that predicted by the EPICUP analyses of the original sample of the cancer of unknown primary.

The EPICUP assay provided a correct histology determination in 174 (96%) of 181 cases of cancer of unknown primary diagnosed by pathological examination under light microscope (table 3). The detailed list of EPICUP versus light microscopy diagnoses in these cases is shown in the appendix (pp 17–20). The EPICUP results were compatible in all 31 cases of cancer of unknown primary in which the comprehensive immunohistochemistry algorithm for diagnosis described in the appendix (p 3) was applied, with the primary site predicted by the battery of used antibodies (table 3; appendix p 21). The tumour type-specific antibodies covered NSCLC (TTF-1), breast (mammoglobin, estogen receptor, and progesterone receptor), liver (HepPar-1), colon (CDX2), bladder (p63), kidney (CD10), and prostate (prostate-specific antigen) tumours, in addition to mesothelioma (calretinin), sarcoma (vimentin), and melanoma (Melan A, S100). Illustrative examples included an interesting case of cancer of unknown primary in a man deemed by EPICUP to have breast cancer; immunohistochemistry analyses revealed the tumour to be positive for mammoglobin, thereby confirming the epigenetic diagnosis.

Of 188 patients with cancer of unknown primary with a diagnosis of primary site origin provided by DNA methylation profiling, overall survival information was available for 114 cases with a median overall survival of 8·1 months (95% CI 0·1–143·4; appendix p 5). The therapy that these 114 patients received is shown in the appendix (pp 22–26). Among these cases, the 92 patients who received chemotherapy (n=84) or radiotherapy (n=8) showed a median overall survival of 9·1 months (95% CI 0·3–57·4). These results are in line with the reported 9-month median overall survival among patients with cancer of unknown primary receiving empirical treatment without considering the tumour of origin.[5] However, the chemotherapy drugs used in these treatments can have a

very different efficacy among the tumour types that give rise to the cancer of unknown primary clinical entity. Thus, we wondered whether those patients with cancer of unknown primary who received a site-specific treatment that fitted the EPICUP prediction showed an improved clinical outcome compared with those who received empirical treatment. We noted that the use of a clinically indicated therapy for the epigenetically predicted tumour type was associated with significantly longer overall survival (n=31) than in those cases who received empirical therapies that did not match the chemosensitive profile of the EPICUP-predicted cancer primary (n=61; HR 3·24, p=0·0051 [95% CI 1·42–7·38]; log-rank p=0·0029; figure part A). The proportional hazards assumption was not violated. The number of deaths was seven (23%) of 31 in the site-specific treatment group and 31 (51%) of 61 in the empiric treatment group (Fisher's test; p=0·013). Examples of site-specific therapies included the use of letrozol, sorafenib, caelyx, 5-fluorouracil, and abraxane in EPICUP-predicted breast, liver, ovary, colon, and pancreatic cancer, respectively (table 4). The Cox multivariable regression model showed that site-dependent therapy was also an independent prognostic factor of overall survival in patients with cancer of unknown primary (figure part B) although sex, age, diagnostic method, histology at diagnosis, number of sites of metastasis, biopsy site, and predicted tumour type were not. Overall, a specific therapy for the DNA methylation-identified original sites conferred a median overall survival of 13·6 months (95% CI 4·1–55·4) compared with 6 months (0·3–57·4) for patients treated with non-specific empirical therapy that did not match the treatment guidelines of the predicted primary tumour (figure part A).

In those patients with cancer of unknown primary in whom the primary tumour was identified as an NSCLC by epigenetic profiling, we screened for *EGFR* (n=16), *ALK* (n=7), *ROS1* (n=6), and *C-MET* (n=7) genetic alterations, which all have an associated clinically approved targeted drug for this tumour type (appendix p 27). In this setting, we detected one *EGFR* mutation, and four *C-MET* gene amplifications. Interestingly, the EPICUP-predicted NSCLC that carried the *EGFR* mutation received treatment with erlotinib, and the patient is alive 55·4 months after the diagnosis, which is unexpected, in view of the usual outcome for patients with cancer of unknown primary. For EPICUP-diagnosed breast tumours (n=12 screened), we found one case harbouring a *HER2* gene amplification and, thus, a cancer of unknown primary amenable to treatment with *HER2* inhibitors. For EPICUP-diagnosed stomach cancers (n=4 screened), none had a *HER2* amplication (appendix p 27).

	Cases (n)	Specific treatments
Breast carcinoma	6	Cyclophosphamide plus doxorubicin plus paclitaxel; capecitabine; denosumab; letrozole
Non-small-cell lung carcinoma	5	Erlotinib; gefitinib; gemcitabine; pemetrexed; vinorelbine
Hepatocellular carcinoma	4	Gemcitabine plus oxaliplatin; bleomycin; iodised oil; sorafenib
Ovarian carcinoma	3	Carboplatin plus paclitaxel; doxorubicin
Endometrial carcinoma	2	Carboplatin plus taxanes; pemetrexed
Colon carcinoma	2	Fluorouracil plus oxaliplatin plus bevacizumab
Mesothelioma	2	Cisplatin plus gemcitabine; pemetrexed
Pancreatic carcinoma	2	Paclitaxel; erlotinib; gemcitabine
Sarcoma	2	Gemcitabine plus docetaxel; ifosfamide; letrozole
Acute lymphoblastic leukaemia	1	Cyclophosphamide plus doxorubicin plus vincristine plus prednisone
Prostate carcinoma	1	Sorafenib
Stomach carcinoma	1	Gemcitabine plus oxaliplatin; capecitabine

EPICUP=microarray DNA methylation signatures.

Table 4: **Cases of cancer of unknown primary classified by tumour types predicted by EPICUP that received specific therapy (n=31)**

Discussion

Our study shows that the use of DNA methylation profiling provides a consistent diagnosis of the primary tumour site in cases of cancer of unknown primary. Furthermore, our results support that the determination of the original tumour type followed by site-specific therapies improves the outcome of these patients compared with those treated empirically, and, in this regard, addresses an unmet need in this area, because only 25% of cases of unknown primary cancer receive a single putative primary tumour diagnosis using light microscopy and immunohistochemical testing.[1,11] In the remaining cases, the immunohistochemical diagnosis is non-specific due to rare altered tissue antigenicity, interobserver and intraobserver variability in interpretation, tissue heterogeneity, the relative insensitivity of the most lineage-specific markers (eg, TTF-1 is very specific, but is only positive in 75–85% of lung adenocarcinomas) and because very often there is no antibody specific for a single tumour type. For example, in cancer of unknown primary with positive staining for CK7 and CK20, immunohistochemistry cannot easily discriminate between pancreatic, gastric, biliary, and ovarian carcinoma, which are tumours that, at an advanced stage, differ subtantially in response to therapy. For those cancers of unknown primary for which a matched antibody exists, such as PSA in prostate cancer and S100 in melanoma, this might represent a clear benefit to the patient. A good example is patients with cancer of unknown primary who are positive for the colorectal marker CDX2; treatment with regimens used for gastrointestinal cancers survived for more than 30 months.[13]

The cancer of unknown primary classifier based on DNA methylation profiles (EPICUP) that we have developed predicted the tissue of origin in 87% of the 216 cases studied. The sample size in our study is similar to that in other studies.[4,12,15] These patients can now receive a less toxic and more site-directed and type-directed therapy that might improve their clinical outcome, as we have observed in our cases and as other

studies have reported.[10–13] We must also consider that the assay was designed to look for similarities between cancers of unknown primary and known primary tumours, not for differences. Consequently, it cannot be excluded that a cancer of unknown primary classified as cancer of a given type by the assay still behaves differently from a typical metastatic cancer of that given type. This can have implications for the administration of primary-specific therapy that might still fail to improve outcome. In this regard, the principle of superiority of primary-specific therapy and the clinical use of the assay should be proven in a prospective cohort or randomised studies.

The test developed here could have implications for the management of cancers of unknown primary, particularly in this new age of medicine in which there is a drive towards more personalised treatments.[26] In the context of cancer of unknown primary, primary site assignment by DNA methylation profile could help to identify about 20% of patients with cancer of unknown primary who have strong responses to systemic or locoregional treatments and longer survival.[27] Importantly, the correct determination of the primary site of origin by EPICUP could guide the screening of drug-actionable mutations. For example in our study, the prediction of an NSCLC or breast primary site can facilitate the development of new molecular tests that reveal *EGFR* mutations or *C-MET* and *HER2* gene amplifications. These patients can now receive a specific targeted treatment and further improve their overall survival. If we search for actionable mutations without knowledge of the precise cellular contex, we might find some unexpected alterations.[26] For example, a *K-RAS* mutation in a patient EPICUP predicts to have an NSCLC with hilar nodes plus brain metastasis might have different clinical implications compared with an EPICUP-diagnosed patient with colorectal cancer with that *KRAS* mutation. Another example would be the discovery of a $BRAF^{V600}$ mutation: if the DNA methylation profile predicts melanoma or thyroid carcinoma, the targeted therapy (BRAF inhibitor) would be more appropriate than if the EPICUP system indicated that the primary site was a colorectal tumour. The observation that patients with cancer of unknown primary who received a tumour type-oriented treatment did better than those receiving non-specific therapy might also be associated with an inherently different prognosis, regardless of the received treatment. Our findings that the tumour type predicted by EPICUP did not significantly affect overall survival, and that treatment type was the only independent prognostic factor do not, however, support this concept; and they instead provide additional reasons to develop tailored therapies for patients with cancer of unknown primary.

The definition of the clinical entity of the cancer of unknown primary is challenging, and as soon as the primary cancer is identified the diagnosis will be changed from cancer of unknown primary to one of the previously occult primary sites. In this regard, the frequency of cancer of unknown primary is probably underestimated.[28] However, the DNA methylation profiler we developed can be extended beyond cancers of unknown primary to other similar clinical conundrums. For example, cases of tumours of "uncertain primary". These include those patients with a previous cancer that subsequently present with metastases that do not match the previous neoplasm; cancers that are unclassifiable due to a poorly differentiated or undifferentiated tumour; and the metastatic cholangiocarcinoma in the presence of an intrahepatic lesion, mimicking a cancer of unknown primary.[29] If these cancers of uncertain primary are added to the cancers of unknown primary studied here, the number of cases that could be assessed approaches 15% of all diagnosed cancers.[28]

One of the strengths of our study is that the diagnoses provided by EPICUP are consistent with the best available knowledge of the clinical, pathological, and molecular features of each case. These include patients in whom the primary tumour was discovered months later, or the cases where the unknown primary was identified by the use of additional antibodies suggested by the EPICUP assay, such as mammoglobin in breast, PSA in prostate, or CDX2 in colon cancer. A final validation of the assay will require extensive and prospective studies of necropsies for patients with cancer of unknown primary; however, it is interesting to note that several investigators believe that the genuine biological entity of cancer of unknown primary is a metastatic tumour for which no primary is identified by any means, including post mortem, or by immunohistochemistry, or imaging. Another advantage of our approach is that it is based on DNA, a material that is stable over time, irrespective of the method of tissue fixation, and that it is not very reactive to change due to minimal external factors, unlike RNA expression levels. In this regard, the assay is likely to cost less than gene expression profiling.[1,14,15] The assay's fast output also favours its further clinical development. Compared with the lengthy diagnostic evaluation process of patients with cancer of unknown primary, a test similar to that described here could possibly provide a diagnosis in 5 days.

In conclusion, our study shows that the use of DNA methylation as a diagnostic instrument for cancers of unknown primary provides an effective means to predict the initially unidentified primary site. This test can also incorporate additional genetic markers to ascertain the best treatment and to avoid morbidity. Although further prospective clinical studies are needed to show its value to increase overall survival of these patients, it is becoming clear that the days of empirical chemotherapy treatment of cancers of unknown primary are reaching their end, and that molecular profiling, such as that described here, will be crucial to the development of tumour type and patient type-specific treatments.

Contributors
SM, AM-C, and ME designed the study, contributed to the analysis, and wrote the first draft of the report. SS, MCdM, HH, and YA provided further data analysis. In-depth patient clinical and pathological characterisation was done by EM, CB, CM, AD-L, PMC, XM-G, CP, JC, DB, LM, DS, RT, JT, and JML. EM, CB, AE-G, GMS, PMC, MR-M, XM-G, RP-C, AA, RL-L, GS, FL, IG, SF, CP, RM, JC, DB, LM, DS, RT, JT, and JML were responsible for patient recruitment. All authors contributed to drafting the work or revising it critically for important intellectual content and made substantial contributions to the concept and design of the study and acquisition, analysis, and interpretation of data.

Declaration of interests
ME reports grants from Ferrer, during the conduct of the study. ME has a patent PCT/EP2012/059687 licensed for Ferrer. AM-C reports personal fees from Ferrer International SA outside of the submitted work. SS reports personal fees from Boehringer Ingelheim Pharma GmbH outside of the submitted work. XM-G reports personal fees from Ferrer International SA outside of the submitted work. JC reports personal fees from Bayer, Johnson & Johnson, Astellas, Amgen, Pfizer, and BMS outside of the submitted work. JT reports grants, personal fees and and non-finacial support of from Amgen, Bayer, Boehringer Ingelheim, Celgene, Chugai, Lilly, MSD, Merck Serono, Novartis, Pfizer, Roche, Sanofi , Symphogen, Taiho, and Takeda outside of the submitted work. JML reports grants, personal fees, non-financial support from Bayer Pharmaceuticals; personal fees, non-financial support from Bristol-Myers Squibb; grants, personal fees, and non-financial support from Boehringer Ingelheim; personal fees from Lilly Pharmaceuticals, Celsion, Biocompatibles, Novartis; and grants, personal fees, and non-financial support from Blueprint Medicines outside of the submitted work. All other authors declare no competing interests.

Acknowledgments
The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme by the ERC Proof-of-Concept Grant EPICUP, under grant agreement No 640696 (to ME); the Cellex Foundation (to ME); the Institute of Health Carlos III (ISCIII) under the Integrated Project of Excellence number PIE13/00022 (ONCOPROFILE); Cancer Research Australia APP1048193 and AP1082604 and Victorian Cancer Agency TRP13062 (to DB, LM, and RT); the Samuel Waxman Foundation (to JML); the Health and Science Deapartments of the Generalitat de Catalunya (to JML and ME); and Ferrer (to ME).

References
1 Varadhachary GR, Raber MN. Cancer of unknown primary site. *N Engl J Med* 2014; **371:** 757–65.
2 Greco FA, Hainsworth JD. Cancer of unknown primary site. In: DeVita VTJ, Hellman S, Rosenberg SA, eds. Cancer: Principles and Practice of Oncology. 9th edn. Philadelphia, PA: Lippincott Williams & Wilkins, 2011. 2033–51.
3 Pavlidis N, Pentheroudakis G. Cancer of unknown primary site. *Lancet* 2012; **379:** 1428–35.
4 Ross JS, Wang K, Gay L, et al. Comprehensive genomic profiling of carcinoma of unknown primary site: new routes to targeted therapies. *JAMA Oncol* 2015; **1:** 40–49.
5 Greco FA, Pavlidis N. Treatment for patients with unknown primary carcinoma and unfavorable prognostic factors. *Semin Oncol* 2009; **36:** 65–74.
6 Massard C, Loriot Y, Fizazi K. Carcinomas of an unknown primary origin—diagnosis and treatment. *Nat Rev Clin Oncol* 2011; **8:** 701–10.
7 Greco FA, Erland JB, Morrissey LH, et al. Carcinoma of unknown primary site: phase II trials with docetaxel plus cisplatin or carboplatin. *Ann Oncol* 2000; **11:** 211–15.
8 Briasoulis E, Kalofonos H, Bafaloukos D, et al. Carboplatin plus paclitaxel in unknown primary carcinoma: a phase II Hellenic Cooperative Oncology Group Study. *J Clin Oncol* 2000; **18:** 3101–07.
9 Culine S, Lortholary A, Voigt JJ, et al. Cisplatin in combination with either gemcitabine or irinotecan in carcinomas of unknown primary site: results of a randomized phase II study—trial for the French Study Group on Carcinomas of Unknown Primary (GEFCAPI 01). *J Clin Oncol* 2003; **21:** 3479–82.
10 Varadhachary GR, Raber MN, Matamoros A, Abbruzzese JL. Carcinoma of unknown primary with a colon-cancer profile-changing paradigm and emerging definitions. *Lancet Oncol* 2008; **9:** 596–99.
11 Varadhachary GR, Spector Y, Abbruzzese JL, et al. Prospective gene signature study using microRNA to identify the tissue of origin in patients with carcinoma of unknown primary. *Clin Cancer Res* 2011; **17:** 4063–70.
12 Hainsworth JD, Rubin MS, Spigel DR, et al. Molecular gene expression profiling to predict the tissue of origin and direct site-specific therapy in patients with carcinoma of unknown primary site: a prospective trial of the Sarah Cannon research institute. *J Clin Oncol* 2013; **31:** 216–23.
13 Varadhachary GR, Karanth S, Qiao W, et al. Carcinoma of unknown primary with gastrointestinal profile: immunohistochemistry and survival data for this favorable subset. *Int J Clin Oncol* 2014; **19:** 479–84.
14 Varadhachary GR, Talantov D, Raber MN, et al. Molecular profiling of carcinoma of unknown primary and correlation with clinical evaluation. *J Clin Oncol* 2008; **26:** 4442–48.
15 Greco FA, Lennington WJ, Spigel DR, Hainsworth JD. Molecular profiling diagnosis in unknown primary cancer: accuracy and ability to complement standard pathology. *J Natl Cancer Inst* 2013; **105:** 782–90.
16 Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* 2003; **349:** 2042–54.
17 Gosden RG, Feinberg AP. Genetics and epigenetics—nature's pen-and-pencil set. *N Engl J Med* 2007; **356:** 731–33.
18 Esteller M. Epigenetics in cancer. *N Engl J Med* 2008; **358:** 1148–59.
19 Esteller M, Garcia-Foncillas J, Andion E, et al. Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N Engl J Med* 2000; **343:** 1350–54.
20 Hegi ME, Diserens AC, Gorlia T, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med* 2005; **352:** 997–1003.
21 Fernandez AF, Assenov Y, Martin-Subero JI, et al. A DNA methylation fingerprint of 1628 human samples. *Genome Res* 2012; **22:** 407–19.
22 Fizazi K, Greco FA, Pavlidis N, et al. Cancers of unknown primary site: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2015; **26**: v133–38.
23 Moran S, Vizoso M, Martinez-Cardús A, et al. Validation of DNA methylation profiling in formalin-fixed paraffin-embedded samples using the Infinium HumanMethylation450 Microarray. *Epigenetics* 2014; **9:** 829–33.
24 Sandoval J, Heyn H, Moran S, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 2011; **6:** 692–702.
25 Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* 2015; **8:** 389–99.
26 Varadhachary G. Carcinoma of unknown primary site: the poster child for personalized medicine? *JAMA Oncol* 2015; **1:** 19–21.
27 Pavlidis N, Petrakis D, Golfinopoulos V, Pentheroudakis G. Long-term survivors among patients with cancer of unknown primary. *Crit Rev Oncol Hematol* 2012; **84:** 85–92.
28 Mnatsakanyan E, Tung WC, Caine B, Smith-Gagen J. Cancer of unknown primary: time trends in incidence, United States. *Cancer Causes Control* 2014; **25:** 747–57.
29 Varadhachary G. New strategies for carcinoma of unknown primary: the role of tissue-of-origin molecular profiling. *Clin Cancer Res* 2013; **19:** 4027–33.