ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

ORIGINAL RESEARCH

Complete genomic characterization in patients with cancer of unknown primary origin in routine diagnostics

L. J. Schipper[1,2†], K. G. Samsom[3†], P. Snaebjornsson[3], T. Battaglia[1], L. J. W. Bosch[3], F. Lalezari[4], P. Priestley[5], C. Shale[5], A. J. van den Broek[6], N. Jacobs[6], P. Roepman[6], J. J. M. van der Hoeven[6], N. Steeghs[7], M. A. Vollebergh[8], S. Marchetti[7], E. Cuppen[2,6,9], G. A. Meijer[3], E. E. Voest[1,2,8] & K. Monkhorst[3*]

[1]Department of Molecular Oncology, Netherlands Cancer Institute, Amsterdam; [2]Center for Molecular Medicine, UMC Utrecht, Utrecht; [3]Department of Pathology, Netherlands Cancer Institute, Amsterdam; [4]Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam; [5]Department of Radiology, Netherlands Cancer Institute, Amsterdam, The Netherlands; [6]Hartwig Medical Foundation Australia, Sydney, Australia; [7]Hartwig Medical Foundation, Amsterdam; [8]Department of Gastroenterology, Netherlands Cancer Institute, Amsterdam; [9]Oncode Institute, Utrecht, The Netherlands

Available online 1 December 2022

Background: In ∼3%-5% of patients with metastatic disease, tumor origin remains unknown despite modern imaging techniques and extensive pathology work-up. With long diagnostic delays and limited and ineffective therapy options, the clinical outcome of patients with cancer of unknown primary (CUP) remains poor. Large-scale genome sequencing studies have revealed that tumor types can be predicted based on distinct patterns of somatic variants and other genomic characteristics. Moreover, actionable genomic events are present in almost half of CUP patients. This study investigated the clinical value of whole genome sequencing (WGS) in terms of primary tumor identification and detection of actionable events, in the routine diagnostic work-up of CUP patients.

Patients and methods: A WGS-based tumor type 'cancer of unknown primary prediction algorithm' (CUPPA) was developed based on previously described principles and validated on a large pan-cancer WGS database of metastatic cancer patients (>4000 samples) and 254 independent patients, respectively. We assessed the clinical value of this prediction algorithm as part of routine WGS-based diagnostic work-up for 72 CUP patients.

Results: CUPPA correctly predicted the primary tumor type in 78% of samples in the independent validation cohort (194/254 patients). High-confidence predictions (>95% precision) were obtained for 162/254 patients (64%). When integrated in the diagnostic work-up of CUP patients, CUPPA could identify a primary tumor type for 49/72 patients (68%). Most common diagnoses included non-small-cell lung ($n = 7$), gastroesophageal ($n = 4$), pancreatic ($n = 4$), and colorectal cancer ($n = 3$). Actionable events with matched therapy options in clinical trials were identified in 47% of patients.

Conclusions: Genome-based tumor type prediction can predict cancer diagnoses with high accuracy when integrated in the routine diagnostic work-up of patients with metastatic cancer. With identification of the primary tumor type in the majority of patients and detection of actionable events, WGS is a valuable diagnostic tool for patients with CUP.

Key words: cancer of unknown primary, whole genome sequencing, tumor type prediction, diagnostic tool

INTRODUCTION

Cancer of unknown primary (CUP) accounts for ∼3%-5% of all metastatic cancers. Despite advancements in the diagnostic arsenal of pathologists and improved imaging modalities over the last decades, primary tumor type remains ambiguous or undetectable for this patient group. Consequently, effective therapeutic options remain very limited. Patients with CUP generally undergo a long diagnostic process, and clinical deterioration prohibits a timely start of treatment in >50% of cases.[1] When CUP-directed treatment is initiated, patients are usually treated with standard platinum-containing chemotherapy (combination) regimens with limited clinical benefit. Consequently, CUP patients continue to have a dismal prognosis, with a median overall survival of 3 months after diagnosis.[2]

Genomic variation has long been recognized as a means for tumor type prediction.[3] For example, the mutational landscape in driver genes across tumor types has been applied as a means to delineate tumor type.[4] Furthermore, mutational profiles of passenger mutations were identified as indicators of cancer etiology, which led to the

*Correspondence to: Dr Kim Monkhorst, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. Tel: +0205122948
E-mail: k.monkhorst@nki.nl (K. Monkhorst).

†Authors contributed equally to this work.

development of widely used COSMIC signatures[5] that can also be used to distinguish tumor types with high accuracy.[6] Next, the topological distribution of driver and passenger mutations varies considerably across cancer types due to chromatin organization and activity differences in the cells of origin.[7] As such, regional mutational density across the genome has diagnostic power.[8] Finally, combinatorial approaches, in which multiple genomic features are grouped within one prediction algorithm, have been shown to further improve classifier accuracy.[8-10] With the increasing use of comprehensive genomic profiling of cancer patients in daily clinical practice, genomics-based tumor type prediction algorithms would be readily implementable in clinical care.

A second advantage to using genomic profiling lies in simultaneous identification of actionable events for targeted treatments. CUP patients have a limited number of therapy options, only consisting of chemotherapy regimens. Almost half of CUP patients, however, harbor an actionable event with direct therapy options in approved, off-label, or clinical trial setting.[11] Clinical efficacy of biomarker-guided targeted therapy based on these actionable events has been reported in several case reports. Moreover, based on a real-world cohort of 3841 CUPs, patients with a targeted therapy approach demonstrated an improved overall survival compared with patients treated with chemotherapy.[12]

Hence, genomic characterization could benefit CUP patients by (i) identification of primary tumor type with subsequent standard of care therapy options and (ii) detection of actionable genomic events to increase the number of therapy options for these patients. For the purpose of primary tumor type prediction in CUPs in the clinical setting, we developed and trained a statistics-based diagnostic tumor tissue of origin prediction tool by combining tumor type-specific drivers, regional mutational density, and mutational profile characteristics on a large pan-cancer whole genome sequencing (WGS) database (>4000 samples). Next, WGS and the cancer of unknown primary prediction algorithm (CUPPA) was applied to 72 CUP patients who received a prospective WGS analysis in a routine diagnostic work-up, to analyze the value of WGS to identify primary tumor type and detect actionable events.

METHODS

Sample collection and WGS procedure

Detailed information on sample collection and WGS procedure can be found elsewhere.[13,14] In summary, samples were collected as part of the CPCT-02 (NCT01855477), DRUP (NCT02925234), and WIDE (NL68609.031.18) clinical trials, approved by medical ethical committees of the University Medical Center Utrecht and the Netherlands Cancer Institute and conducted in concordance with the Declaration of Helsinki, Dutch law, and Good Clinical Practice. Fresh tumor samples were used for DNA isolation and sequenced at 90-100× coverage using uniform sample and data processing procedures by the Hartwig Medical Foundation (Hartwig).[15] A 10 ml blood withdrawal was used to perform germline sequencing at 30× coverage to allow for somatic variant calling and excluding germline variation. After sequencing, genomic and clinical information including primary tumor type was stored in the Hartwig database. Samples with unknown or undocumented primary tumor type were excluded for development and validation of the tumor type prediction algorithm.

WGS-based tumor type prediction model

We developed CUPPA, a statistical model that weighs multiple genomic features, to find resemblance of a sample compared with different cohorts of samples based on their primary tumor origin (reference cohorts) (Fig. 1). In case of a limited number of distinct samples of certain tumor origin, different primary tumor types were grouped into a single reference cohort based on clinicopathological similarities (Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2022.100611). Samples not fitting any reference cohort were excluded from analysis. Three orthogonal DNA classifiers, each with predictive power for tissue of origin, were combined to reach an overall prediction. In more detail, independent classifiers for positional mutational distribution, relative contextual single nucleotide variants used for COSMIC signatures, and presence of cancer type-specific drivers and passenger mutational features, and the combined classifier are assigned a relative similarity likelihood to each primary cancer origin cohort with the sum of the likelihoods adding up to 1 across the 29 reference cohorts. The similarity likelihood is provided with every CUPPA prediction as 'prediction likelihood'. Samples derived from males are excluded from matching 'Ovary' and 'Uterus' cancer reference cohorts and samples from females are excluded from matching the 'Prostate' reference cohorts in the combined classifier. A detailed description on the calculation of all classifiers can be found in the Supplementary methods, available at https://doi.org/10.1016/j.esmoop.2022.100611.

Internal validation

At time of analysis, the Hartwig database consisted of 4509 samples with known histopathological-based primary origin. These samples were randomly divided in a reference set (90%, $n = 4058$) and a test set (10%, $n = 451$). CUPPA was applied to the test set with the reference set as base for reference cohort determination. CUPPA overall prediction was compared with known histopathology-based primary origin to determine predictive performance on tumors with known origin. Similarity likelihood scores were used to determine a cut-off for high-confidence and low-confidence CUPPA predictions.

Independent validation cohort

In the period January 2021-September 2021, patients underwent WGS analysis as part of their regular diagnostic work-up at the Netherlands Cancer Institute. WGS was either indicated for identification of therapy options, or for supporting classification of diagnostically challenging

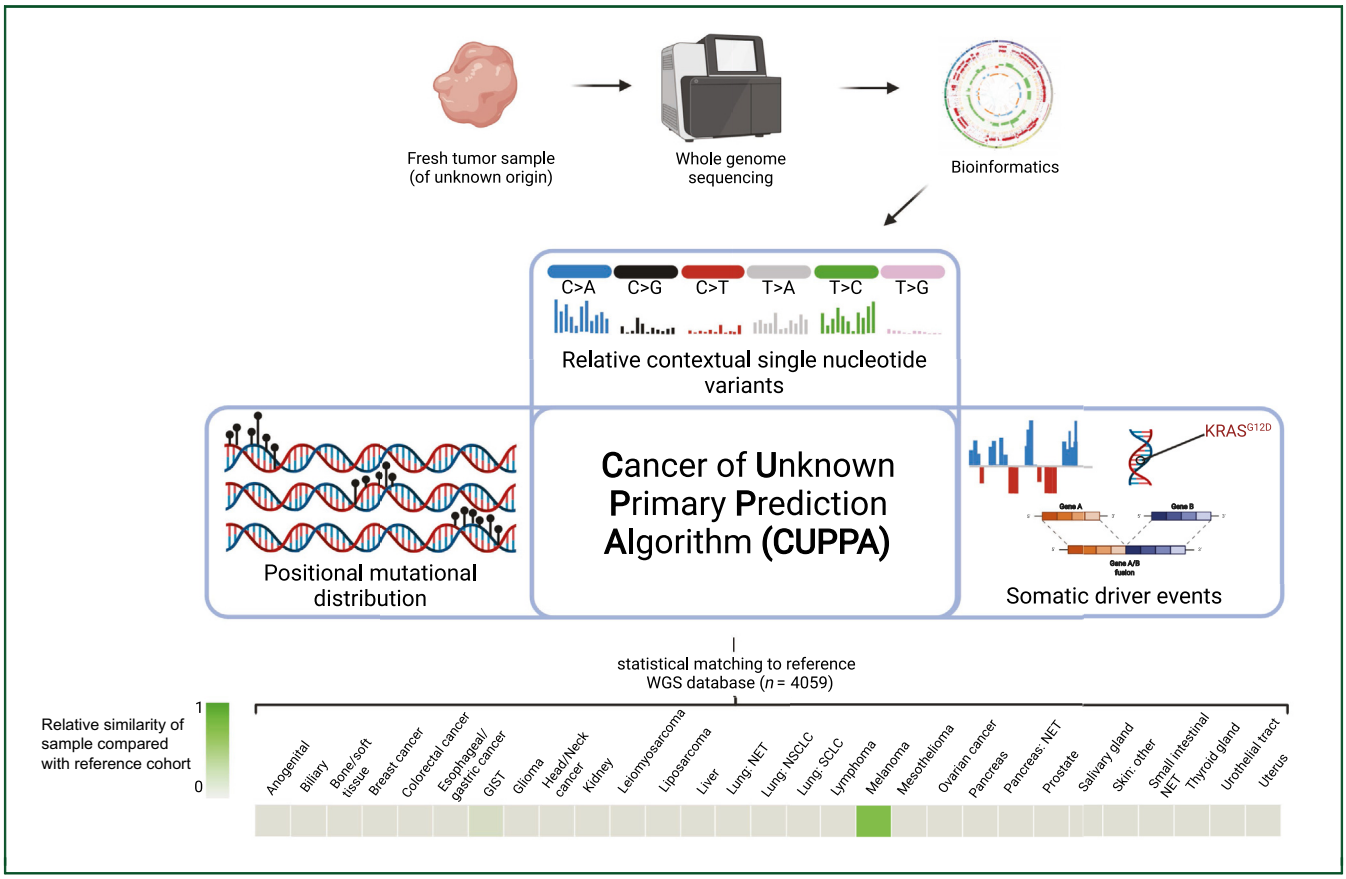

Figure 1. The cancer of unknown primary prediction algorithm (CUPPA). Created with BioRender.com. GIST, gastrointestinal stromal tumor; NET, neuroendocrine tumor; NSCLC, non-small-cell lung cancer; WGS, whole genome sequencing

tumors or CUPs. All WGS analyses (data generation, data processing, actionable variant, and CUPPA reporting) were carried out in a centralized facility, operated by Hartwig. Use of data for this study was approved by the Institutional Review Board of the Netherlands Cancer Institute. All patients provided written informed consent for use of data for research purposes. Data were anonymized and handled in accordance with the Code for Proper Secondary Use of Human Tissue in the Netherlands.

Value in clinical setting

Treating oncologists and/or pathologists had the opportunity to request a CUPPA analysis for CUP patients and otherwise diagnostically complex tumors. Three patients have been described elsewhere.[16] CUPs were defined as tumors with unknown origin or histological type. CUPPA was tested prospectively in all patients in the independent validation cohort. In a dedicated research meeting, with participation of expert pathologists and medical oncologists, CUPPA analyses were reviewed. Where needed, additional diagnostic tests were carried out to adequately interpret clinical characteristics. For each CUPPA prediction, it was determined whether the prediction corresponded with the differential diagnosis of the pathologist before WGS analysis, or whether the prediction did not fit the clinical presentation of the patient.

Actionability

Previously described oncogenic driver likelihood scores were used to assess pathogenicity of variants, and if needed, additional diagnostic analyses were carried out to determine gene and/or protein expression.[13] For every oncogenic variant, it was determined whether the variant could elicit clinical trial participation in an ongoing trial within any Dutch hospital at time of the WGS analysis. For this, variants were correlated with the iClusion database, in which all clinical trials (phase I-III) running in the Netherlands and their eligibility criteria, including genomic indications, are stored. In case genomic variants were regarded as actionable at the time of WGS analysis, but compelling evidence for inefficacy led to discontinuation of the associated clinical trial (cohort) at a later time, these variants were retrospectively disregarded as viable clinical trial options.

Model evaluation

For the multiclass metastatic tumor origin classification task, an initial evaluation was carried out using conventional definitions of sensitivity, specificity, F1-score, and balanced accuracy on the argmax of the class probabilities (maximum cancer type probability). We next assessed the complete model predictions using a one-versus-rest strategy to generate a single binary classification problem per class and computed the mean area under the curve-receiver operator

curve (AUC-ROC) and mean average precision (AP) through micro-, macro-, and weighted-averaging. Top-k accuracy estimates for k ∈ 1, 3, 5 in low-confidence tumor type predictions (no probability >0.80), was computed by determining how often the true tumor type classification was included in the top-k model predictions, normalized by the number of samples.

RESULTS

Model performance

Overall, CUPPA was able to correctly classify 84% and 78% of samples in the internal (n = 451) and independent validation cohort (n = 254), respectively (Figure 1A, multimedia component 3) (Figure 2A, Supplementary Figure S1, available at https://doi.org/10.1016/j.esmoop.2022.100611). Micro-averaged one-versus-rest ROCs showed an AUC of 0.993 and 0.986, and APs were 0.993 and 0.879 for the internal and independent validation cohort, respectively (Figure 2B and C). Predictive performance of CUPPA was higher for common tumor types with a larger number of samples in the corresponding reference cohort (Supplementary Figures S1 and S2, available at https://doi.org/10.1016/j.esmoop.2022.100611).

To achieve a predictive precision of 95% in the internal validation cohort, we used a similarity likelihood score of 0.8 as a cut-off for high-confidence predictions (Figure 2D). Within the internal validation cohort, 75% of samples (n = 340/451) had a probability score of ≥0.8 (Supplementary Figure S3, available at https://doi.org/10.1016/j.esmoop.

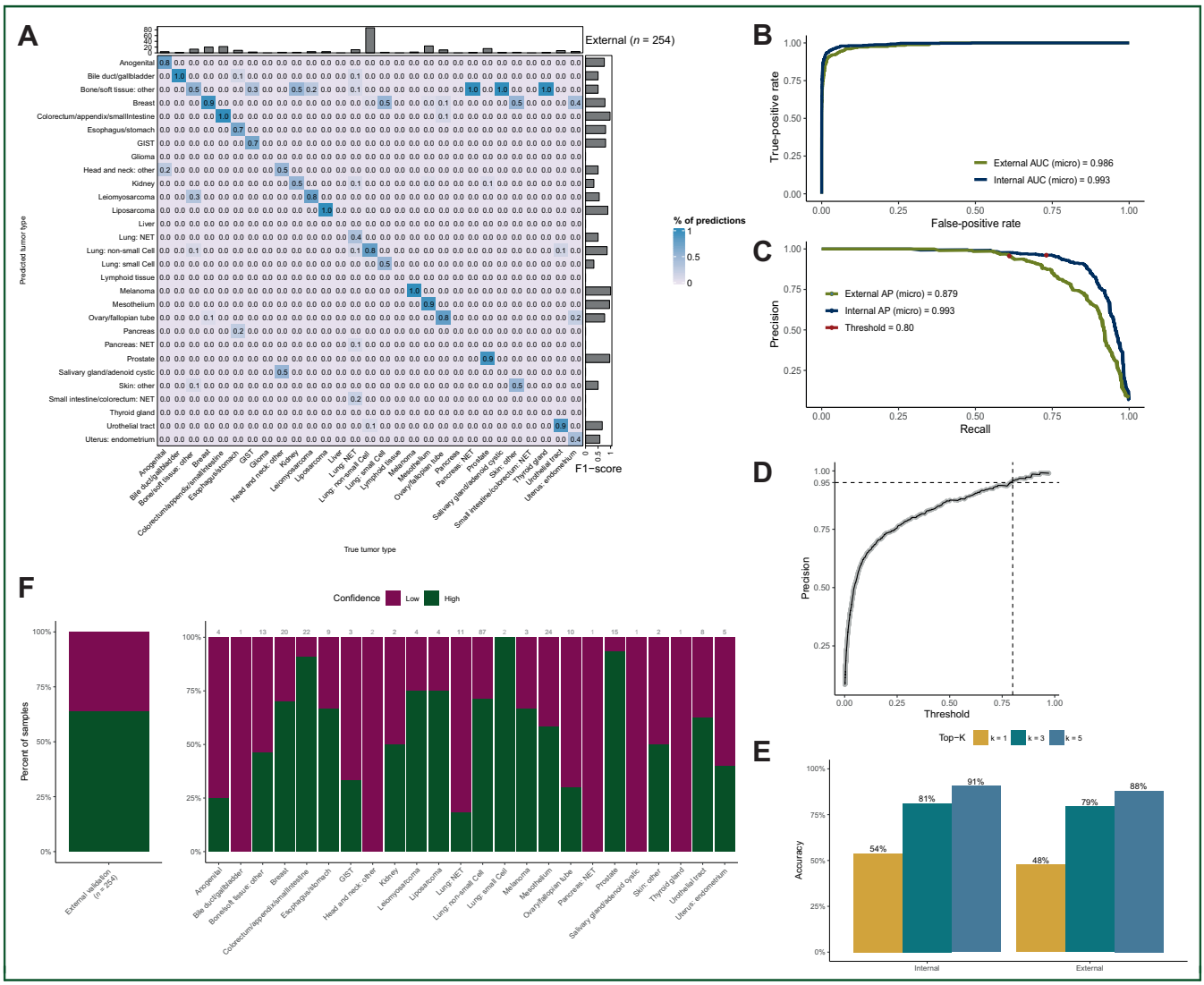

Figure 2. Predictive performance of CUPPA. (A) Confusion matrix showing model performance across 29 tumor types in the external validation cohort. The confusion matrices for the reference set and internal validation set can be found in the supplementary figures (Supplementary Figures S1 and S2). (B) Receiver operator curves and precision-recall curves (C) showing the overall model performance. (D) By using the probability score generated by the model as a cut-off, a high predictive precision (95%) was reached at a score of 0.8. The corresponding metrics at this cut-off are plotted in panel B and C. (E) In low-confidence predictions (<0.8), predictive precision was lower. The high top-3 and top-5 model accuracy demonstrated, however, that low-confidence predictions can be used as a mean to derive a differential diagnosis to correlate with clinicopathological differential diagnoses. (F) In total, 64% of samples reached a high-confidence prediction. Distribution of high- and low-confidence predictions varied across tumor types.

AP, average precision; AUC, area under the curve; CUPPA, cancer of unknown primary prediction algorithm; GIST, gastrointestinal stromal tumor; NET, neuroendocrine tumor.

2022.100611). At this cut-off, taking samples with a high-confidence prediction only, micro-averaged true-positive rate and false-positive rate were 0.961 and 0.0013, respectively. In the independent validation cohort, a lower percentage of samples reached a high-confidence prediction (162/254, 64%), possibly a result of the relative distribution of tumor types (Figure 2F, Supplementary Table S2 available at https://doi.org/10.1016/j.esmoop.2022.100611). Subsequently, the lower overall precision for the external compared with the internal validation cohort (78% versus 84%) could be contributed to tumor type distribution. All rare cancer samples without a matching reference cohort received a low-confidence prediction (Supplementary Table S3, available at https://doi.org/10.1016/j.esmoop.2022.100611).

In samples with a low-confidence prediction (probability score <0.8), predictive precision was 48% (Figure 2E). To determine the value of low-confidence predictions in narrowing down differential diagnoses, we evaluated top-k accuracy, i.e. the predictive accuracy of the model taking the k highest confidence predictions. CUPPA reached a top-3 accuracy of 79% and a top-5 accuracy of 88% in low-confidence predictions, indicating that top-5 predictions can serve as a genome-based differential diagnosis to correlate with the clinical, histopathological, and immunohistochemistry-derived differential diagnosis.

Clinical value in routine diagnostic work-up

The external validation cohort comprised of 254 patients with a variety of known tumor types that underwent a WGS analysis and prospective CUPPA analysis as part of their regular diagnostic work-up (Supplementary Figure S3, available at https://doi.org/10.1016/j.esmoop.2022.100611). High-confidence predictions of CUPPA (64%) were prospectively incorporated within the diagnostic work-up and correlated with the clinical presentation of patients. Initially, 7 of 162 (4.3%) high-confidence predictions did not match the presumed tumor type at time of diagnosis. Among these seven cases, three diagnostic revisions were made based on WGS and CUPPA analysis. This included one patient with a revision from a large-cell neuroendocrine carcinoma to a grade 3 neuroendocrine tumor of the lung, which was also more consistent with the clinical presentation of slow tumor progression. For a second patient, the diagnosis was revised from a sarcomatoid carcinoma of the lung to an undifferentiated sarcoma. Finally, one patient was diagnosed with small-cell lung cancer with widespread metastases in lymph nodes, pancreas, liver, skin, breast, and bones. Based on a high-confidence prediction for breast cancer, the lesion within the breast was regarded as the primary tumor, which fitted better with the clinical presentation: a young, female, non-smoking patient.

For two other discordant patients, CUPPA prediction did match with the expected tumor biology. First, a patient was diagnosed with a mature teratoma of the ovary with differentiation into intestinal type adenocarcinoma (Supplementary Figure S4, available at https://doi.org/10.1016/j.esmoop.2022.100611). Consistent with tumor biology, this tumor was classified by CUPPA as a colorectal/small intestinal cancer sample with a high prediction (0.82). Second, a patient with non-small-cell lung cancer (NSCLC) was biopsied after progression on chemoimmunotherapy. Although the case was initially not regarded as SCLC on morphological grounds, small-cell transformation of NSCLC after immunotherapy has previously been associated with TP53 and RB1 loss.[17]

When excluding these five cases, overall accuracy of CUPPA in high-confidence predictions was 155/157 (98.7%). Two cases with a high-confidence prediction were definitively misclassified; this includes a patient with diffuse-type gastric carcinoma that was classified as a pancreatic carcinoma and an undifferentiated pleiomorphic sarcoma that was classified as a leiomyosarcoma. Misclassifications of sarcoma subtypes are a known pitfall of CUPPA (Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2022.100611).

Tumor type predictions in CUP patients

Next, we applied WGS and CUPPA to 72 patients referred to the Netherlands Cancer Institute with a clinical diagnosis of CUP, for whom extensive pathological, radiological, and endoscopic modalities failed to identify a primary tumor type (Table 1, multimedia component 9, multimedia component 10) (Table 1, Supplementary Tables S4 and S5, available at https://doi.org/10.1016/j.esmoop.2022.100611). WGS was generally carried out early in the disease course. A total of 26 of these patients had a history of previous malignancy. WGS results and CUPPA predictions were correlated with the differential diagnosis that was composed based on clinical and pathological characteristics. For 37 patients (51%), CUPPA was able to provide a high-confidence prediction (probability score >0.8, Figure 3A). In all of these patients, the high-prediction score was consistent with one of the tumor types considered in the differential diagnosis before WGS analysis. For 35 of the 37 patients, the high-confidence prediction led to a definitive diagnosis. The most common diagnoses included non-small-cell lung cancer ($n = 7$), gastroesophageal cancer ($n = 4$), pancreatic cancer ($n = 4$), and colorectal/small intestinal cancer ($n = 3$). For 2 of the 35 patients, the high-confidence tumor type prediction did not add to the clinicopathological considerations before WGS/CUPPA analysis. Based on clinicopathological diagnostics, these two patients were diagnosed with a squamous cell CUP, most likely from the anogenital area. Since anogenital carcinomas are grouped within one prediction category, these two high-confidence predictions (probability score of 98.4% and 98.3%, respectively) did not add any information regarding the primary site within the anogenital area. Of note, squamous cell carcinomas of the anogenital area are generally treated with platinum-containing chemotherapy regimens.

In 35 CUP patients, a low-confidence prediction was reached (Figure 3B). By correlating the low-confidence predictions with the clinicopathological differential diagnosis of

Table 1. Baseline characteristics of 72 CUP patients

	Total cohort ($n = 72$)
Age at WGS analysis (years) median (range)	62 (18-81)
Sex *n* (%)	
Female	31 (43)
Male	41 (57)
Disease duration in months median (range)	2 (0-37)
Disease stage *n* (%)	
Metastatic	67 (93)
Locally advanced	5 (7)
Number of previous systemic therapy lines *n* (%)	
0	60
1	9
2	2
3	1
Oncological history *n* (%)	26 (36)
Tumor localization *n* (%)	
Lymph nodes	55 (76)
Lung	25 (35)
Liver	24 (33)
Bone	18 (25)
Peritoneum	14 (19)
Adrenal glands	8 (11)
Skin/subcutis	5 (7)
Other	15 (21)
Morphology[a] *n* (%)	
Adenocarcinoma	32 (44)
Adenosquamous carcinoma	1 (1)
Squamous cell carcinoma	10 (14)
Neuroendocrine carcinoma	2 (3)
Undifferentiated malignancy	21 (29)
Other	6 (8)

[a] Morphology as defined before reaching a diagnosis.

patients, however, these predictions proved to be informative in 12 patients where the differential diagnosis was narrowed down to a probable diagnosis as assessed by a panel of expert pathologists. A description on the diagnostic considerations and the added value of low-confidence predictions for these cases can be found in Supplementary Table S6, available at https://doi.org/10.1016/j.esmoop.2022.100611. In two additional patients with a low-confidence prediction, WGS detected a disease-defining genomic event (Merkel cell polyoma virus and a SS18-SSX1 fusion). Taken together, with high-confidence ($n = 35$) and low-confidence ($n = 12$) CUPPA predictions and detection of diagnostic biomarkers ($n = 2$), WGS was able to establish a diagnosis within the clinical context for 49/72 (68%) CUP patients.

Therapeutic opportunities

With the identification of a primary tumor type, CUP-directed chemotherapy regimens can be substituted with tumor type-specific therapy regimens. In addition, genomic characterization allows for the detection of actionable variants, leading to biomarker-based therapeutic opportunities in clinical trial setting. We assessed genome-based actionability by assessing the number of experimental therapy options in ongoing Dutch clinical trials. For 34/72 CUP patients (47%), one or more therapy options were identified (Figure 4A, Supplementary Table S7, available at https://doi.org/10.1016/j.esmoop.2022.100611). In 10 patients, multiple therapy options were identified, allowing for subsequent therapies or providing a rationale for combinatorial therapies. In most cases, this consisted of combinations with checkpoint inhibitors (Figure 4B). Actionable events were identified in 7/21 patients with a definite CUP (33%) for whom tumor type-directed therapy remained unavailable (Figure 4A). Taken together, adding up tumor type-directed regular therapy and detection of biomarker-based therapy options, WGS had potential therapeutic implications in the majority of CUP patients (56/72 patients, 78%).

DISCUSSION

We developed a tumor type prediction algorithm based on previously described genomic predictors that can be integrated in the regular diagnostic work-up with a short turnaround time of <2 weeks. As such, complete genomic characterization of CUP led to identification of a primary tumor type and detection of actionable events in 68% and 47% of patients ($n = 72$), respectively. In all high-confidence predictions, the prediction was consistent with the clinicopathological differential diagnoses that were proposed before WGS analysis. Moreover, by integrating WGS with the clinicopathological differential diagnosis based on regular diagnostic tests, low-confidence predictions led to a diagnosis in 34% of patients. In our experience, clinical utility of CUPPA was optimal when used as an addition to the extensive diagnostic work-up that CUP patients generally receive, rather than using genome-based tumor type prediction as a stand-alone test. As such, CUPPA also holds the potential to recognize misdiagnoses, as 3/162 patients with a known diagnosis received a diagnostic revision after WGS analysis.

With this study, we focused on improving the diagnostic work-up among patients with CUP, both by identifying the primary tumor type and through identification of additional therapy options. The impact of providing more treatment opportunities on overall survival of these patients is beyond the scope of this study. CUPs are regarded as tumors that are relatively refractory to systemic treatment, and it remains to be seen whether tumor type-directed (chemo)therapy will benefit these patients in terms of survival. In a prospective clinical trial, 194 patients received a tumor type directed therapy based on a molecular gene expression classifier. Median overall survival time was higher compared with a historical cohort of CUP patients receiving CUP-directed chemotherapy (12.5 versus 9.1 months).[18] In contrast, in two randomized clinical trials, site-specific treatment failed to improve median overall or progression-free survival compared with empirical chemotherapy.[19,20] Results on efficacy of targeted therapy in this group have been more promising so far. Several case reports have demonstrated clinical benefit of molecularly guided targeted therapy, with durable partial responses or disease stabilization.[21-25] Intriguingly, based on real-world data of 3841 CUP patients, a targeted therapy demonstrated improved survival outcomes compared with regular chemotherapy. Nonetheless, conclusive clinical evidence on the efficacy of targeted

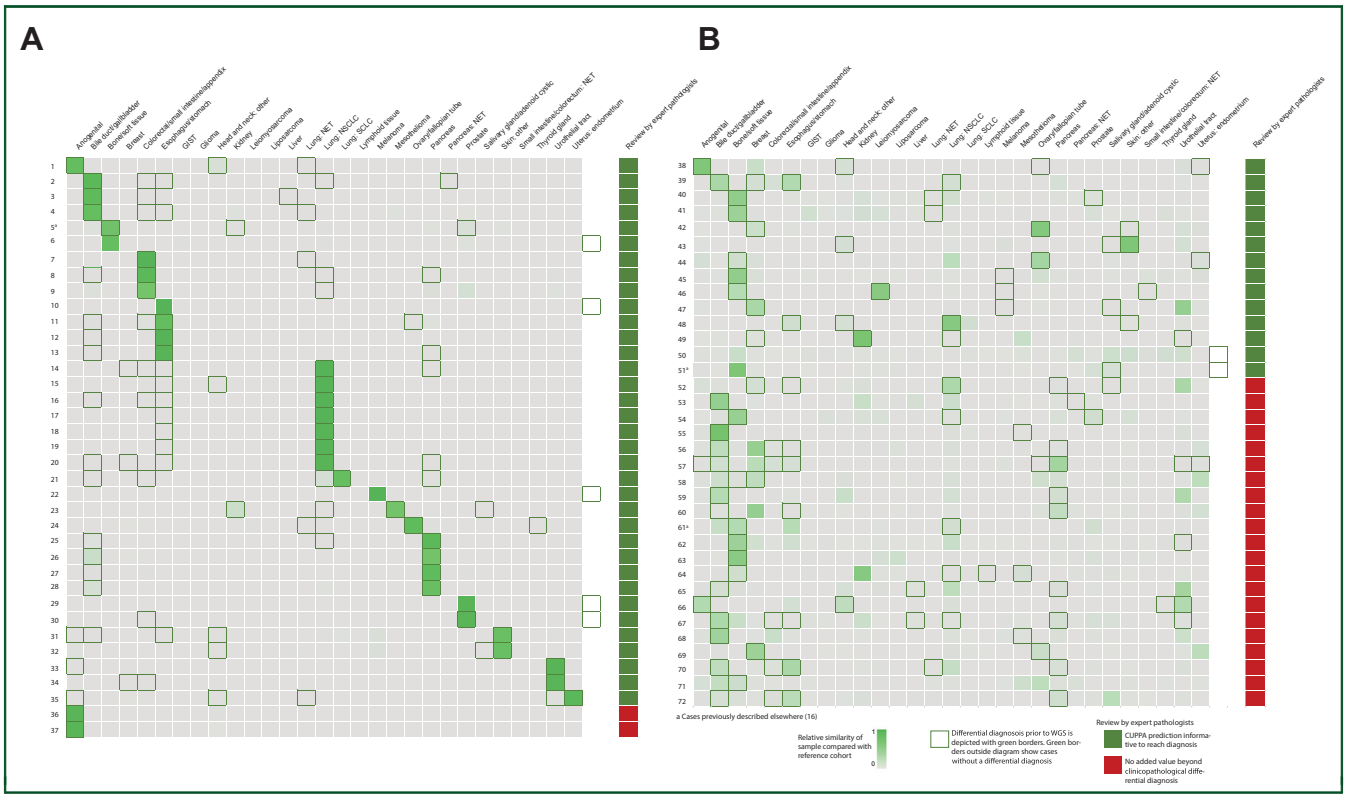

Figure 3. Application of CUPPA in diagnostic work-up of 72 CUP patients.
For 37 (51%) patients, a high-confidence prediction was reached (panel A). All high-confidence predictions were consistent with the differential diagnosis (green borders) before WGS. For 35/37 cases, a final diagnosis could be reached (dark green boxes under 'Review by expert pathologists'). For two patients, the prediction did not provide additional information for the diagnosis (red boxes). For the remaining 35 cases, a low-confidence prediction was reached (panel B). When integrated with prior clinicopathological differential diagnoses, this prediction proved to be informative to reach a diagnosis in 12 patients (patients 38-49). In two additional patients (50 and 51), a disease-defining genomic event was detected with WGS. A description of the clinical value of CUPPA in low-confidence predictions can be found in Supplementary Table S6. the Supplementary material, available at https://doi.org/10.1016/j.esmoop.2022.100611.
CUPPA, cancer of unknown primary prediction algorithm; GIST, gastrointestinal stromal tumor; NET, neuroendocrine tumor; NSCLC, non-small-cell lung cancer; WGS, whole genome sequencing.

therapies in CUP patients is currently not available. To assess this systematically, the international multicenter CUPISCO trial is currently ongoing in which CUP patients are randomized (3 : 1) after three cycles of standard first-line chemotherapy and subsequently allocated to matched targeted therapy.[26]

Of note, our approach has one major direct clinical advantage. All relevant diagnostic tests, including WGS

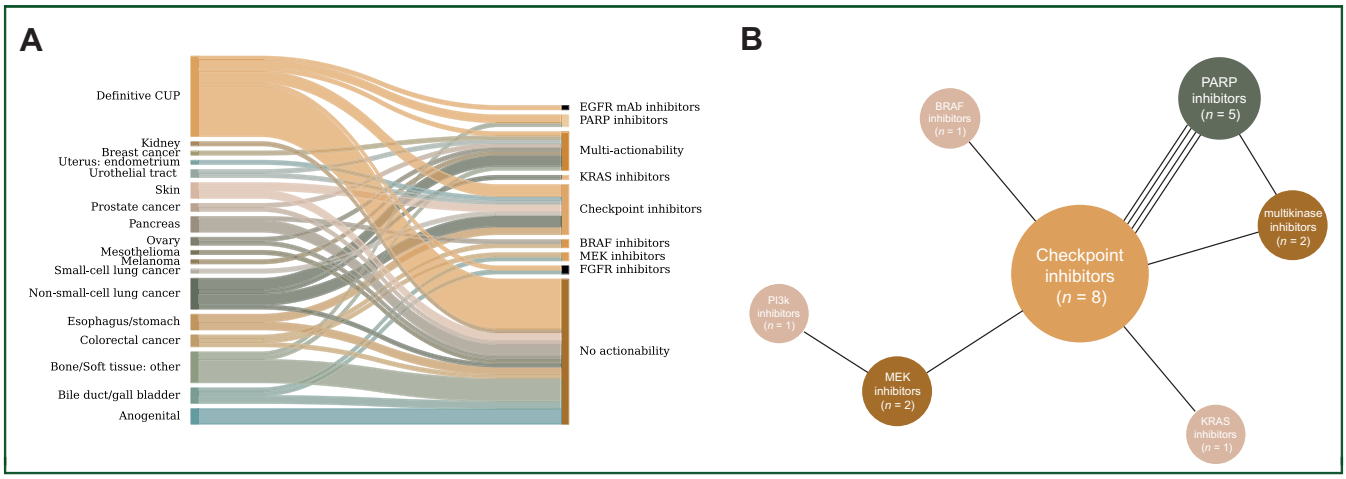

Figure 4. Biomarker-based therapy options detected with WGS in CUP patients. In 47% of patients, an actionable event was identified (panel A). In patients with a definitive CUP ($n = 21$), an actionable event was identified in 33% (7 patients). For 10 patients, multiple therapy options were identified. In panel B, each line represents 1 of these 10 patients, showing the multiple therapy options identified in each patient.
CUP, cancer of unknown primary; WGS, whole genome sequencing.

analysis, were immediately requested at first suspicion of a CUP. WGS reports were routinely delivered within 2 weeks for the prospective part of this study.[27] As a result, the time to diagnosis is shortened and a patient's tumor can be classified according to origin or as a definitive CUP. Shortening the diagnostic work-up could potentially improve clinical outcomes, as patients start systemic therapy with a better clinical condition. More importantly, the uncertainty regarding diagnosis poses a high psychological burden on patients.

Our tumor prediction algorithm is a statistical model that compares multiple genomic features of the sample of interest with a large reference database. An inherent limitation to this approach lies in the dependency on adequate sample size per reference cohort. Accurate tumor type classification is more challenging for cancer types with a limited number of samples in the reference cohort and low-confidence predictions should be interpreted with caution and regard for current pitfalls of the algorithm (Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2022.100611). Also, rare cancers are inevitably misclassified by the prediction algorithm, as they cannot be allocated to non-existing reference cohorts. To address this problem, our overarching ambition is to develop a learning health care system. Any new sample of known tumor origin that is being sequenced at Hartwig is automatically added to the reference database, providing patients have given consent for re-use of their data. Likewise, other WGS datasets of (metastatic) cancer patients generated outside Hartwig can be adjoined. Furthermore, enrichment strategies, in which samples of specific tumor types are sequenced and added to the reference database, can be implemented to reach a minimum threshold of samples for rare cancers or cancer types that are most relevant for differential and CUP diagnosis. Moreover, future discoveries in tumor type-specific genomic disparities can be added in the algorithm as new predictors. For example, non-coding somatic drivers[28] or microbiome analyses[29] are possible with already available sequencing data. Finally, other modalities, like gene expression,[30] methylation profiles,[31] or digital whole-slide images,[32] can be easily incorporated within the algorithm itself, although such strategies require additional analyses of the complete reference database and may be partially redundant with each other and/or genomic features. The clinical value of incorporating multiple modalities has been demonstrated in a recently published cohort of 70 CUP patients, in which combined genomic, transcriptomic, and methylome revealed a probable tumor type in 89% and treatment recommendations in 80% of patients.[33] With the ability to continuously optimize the prediction algorithm, and increase the number of samples within the reference database, predictive accuracy of CUPPA is likely to improve over time. Adding other modalities will very likely improve the algorithm further.

In conclusion, complete genomic characterization with WGS was demonstrated to have a significant added value to the diagnostic arsenal for CUP patients. By integrating WGS into the diagnostic work-up, a primary tumor could be determined in 68% of CUP patients, and actionable events for matched therapy decisions in 47% of patients. Follow-up research on the efficacy of site-specific and matched targeted therapies is conducted internationally. Regardless, shortening of the diagnostic work-up allows for earlier treatment initiation, reduced diagnostic work-up, and limited duration of CUP-associated psychological burden. With these considerations in mind, WGS is now being reimbursed for CUP patients in the Netherlands.

FUNDING

This work was supported by ZonMw, the Netherlands Organization for Health Research and Development (Project Number 446002004) as part of the WIDE study, including an in-kind contribution of Hartwig Medical Foundation (no grant number). The study protocol has been independently peer-reviewed by ZonMw. ZonMw had no role in the design nor the collection, analysis, and interpretation of the data, nor the writing of the manuscript.

DISCLOSURE

PS has done unrelated consultancy for Merck Sharp & Dohme (MSD) and Bayer and received payment from MEDtalks for educational presentation. EC reports consultancy fees and support for attending meetings and travelling from Illumina. EEV is member of the supervisory board of Hartwig. GAM is co-founder and board member (CSO) of CRCbioscreen BV, he has a research collaboration with CZ Health Insurances (cash matching to ZonMW grant), and he has research collaborations with Exact Sciences, Sysmex, Sentinel Ch. SpA, Personal Genome Diagnostics (PGDX), DELFi; these companies provide materials, equipment and/or sample/genomic analyses. He is an advisory board member of 'Missie Tumor Onbekend'. KM reports research grants from AstraZeneca and speaker's fees from MSD, Roche, AstraZeneca and Benecke. KM received consultancy fees from Pfizer, Bristol Myers Squibb, Roche, MSD, AbbVie, AstraZeneca, Diaceutics, Lilly, Bayer, Boehringer Ingelheim and non-financial support from Roche, Takeda, Pfizer, PGDx, and DELFi. All other authors have declared no conflicts of interest.

REFERENCES

1. Meijer L, Verhoeven RHA, de Hingh IHJT, et al. Extensive diagnostic work-up for patients with carcinoma of unknown primary. *Clin Exp Metastasis*. 2021;38(2):231-238.
2. Hemminki K, Bevier M, Hemminki A, Sundquist J. Survival in cancer of unknown primary site: population-based analysis by site and histology. *Ann Oncol*. 2012;23(7):1854-1863.
3. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999;286(5439):531-537.
4. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013;502(7471):333-339.
5. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415-421.

6. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94-101.
7. Polak P, Karlić R, Koren A, et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature*. 2015;518(7539): 360-364.
8. Jiao W, Atwal G, Polak P, et al. A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns. *Nat Commun*. 2020;11(1):728.
9. Chakraborty S, Martin A, Guan Z, Begg CB, Shen R. Mining mutation contexts across the cancer genome to map tumor site of origin. *Nat Commun*. 2021;12(1):3051.
10. Nguyen L, van Hoeck A, Cuppen E. Machine learning-based tissue of origin classification for cancer of unknown primary diagnostics using genome-wide mutation features. *bioRxiv*. 2021. https://doi.org/10.1101/2021.10.05.463244.
11. Lombardo R, Tosi F, Nocerino A, et al. The quest for improving treatment of cancer of unknown primary (CUP) through molecularly-driven treatments: a systematic review. *Front Oncol*. 2020;10:533.
12. Park C, Georlette D, Korn WM, et al. 1139P Carcinoma of unknown primary (CUP): the role of tumor genomic profiling. *Ann Oncol*. 2021;32:S926.
13. Priestley P, Baber J, Lolkema MP, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature*. 2019;575(7781):210-216.
14. Samsom KG, Bosch LJW, Schipper LJ, et al. Study protocol: whole genome sequencing implementation in standard diagnostics for every cancer patient (WIDE). *BMC Med Genomics*. 2020;13(1):169.
15. Roepman P, de Bruijn E, van Lieshout S, et al. Clinical validation of whole genome sequencing for cancer diagnostics. *J Mol Diagn*. 2021;23(7):816-833.
16. Schipper LJ, Monkhorst K, Samsom KG, et al. Clinical impact of prospective whole genome sequencing in sarcoma patients. *Cancers (Basel)*. 2022;14(2):436.
17. Arakawa S, Yoshida T, Shirasawa M, et al. RB1 loss induced small cell lung cancer transformation as acquired resistance to pembrolizumab in an advanced NSCLC patient. *Lung Cancer*. 2021;151:101-103.
18. 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*. 2012;31(2):217-223.
19. Hayashi H, Kurata T, Takiguchi Y, et al. 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):570-579.
20. Fizazi K, Maillard A, Penel N, et al. A phase III trial of empiric chemotherapy with cisplatin and gemcitabine or systemic treatment tailored by molecular gene expression analysis in patients with carcinomas of an unknown primary (CUP) site (GEFCAPI 04). *Ann Oncol*. 2019;30:v851.
21. Palma NA, Ali SM, O'Connor J, et al. Durable response to crizotinib in a MET-amplified, KRAS-mutated carcinoma of unknown primary. *Case Rep Oncol*. 2014;7(2):503-508.
22. Kato S, Krishnamurthy N, Banks KC, et al. Utility of genomic analysis in circulating tumor DNA from patients with carcinoma of unknown primary. *Cancer Res*. 2017;77(16):4238.
23. Subbiah IM, Tsimberidou A, Subbiah V, Janku F, Roy-Chowdhuri S, Hong DS. Next generation sequencing of carcinoma of unknown primary reveals novel combinatorial strategies in a heterogeneous mutational landscape. *Oncoscience*. 2017;4(5-6):47-56.
24. Tan DSW, Montoya J, Ng Q-S, et al. Molecular profiling for druggable genetic abnormalities in carcinoma of unknown primary. *J Clin Oncol*. 2013;31(14):e237-e239.
25. Zhao P, Peng L, Wu W, et al. Carcinoma of unknown primary with EML4-ALK fusion response to ALK inhibitors. *Oncologist*. 2019;24(4): 449-454.
26. Pauli C, Bochtler T, Mileshkin L, et al. A challenging task: identifying patients with cancer of unknown primary (CUP) according to ESMO guidelines: the CUPISCO trial experience. *Oncologist*. 2021;26(5):e769-e779.
27. Samsom K, Monkhorst K, Schipper LJ, et al. Feasibility of whole-genome sequencing in routine clinical practice. *J Clin Oncol*. 2021;39(suppl 15):3013.
28. Rheinbay E, Nielsen MM, Abascal F, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature*. 2020;578(7793): 102-111.
29. Poore GD, Kopylova E, Zhu Q, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature*. 2020;579(7800): 567-574.
30. Penson A, Camacho N, Zheng Y, et al. Development of genome-derived tumor type prediction to inform clinical cancer care. *JAMA Oncol*. 2020;6(1):84-91.
31. Koelsche C, Schrimpf D, Stichel D, et al. Sarcoma classification by DNA methylation profiling. *Nat Commun*. 2021;12(1):498.
32. Lu MY, Chen TY, Williamson DFK, et al. AI-based pathology predicts origins for cancers of unknown primary. *Nature*. 2021;594(7861):106-110.
33. Möhrmann L, Werner M, Oleś M, et al. Comprehensive genomic and epigenomic analysis in cancer of unknown primary guides molecularly-informed therapies despite heterogeneity. *Nat Commun*. 2022;13(1): 4485.