

## **Molecular diagnosis and prognosis of cancers of unknown-primary (CUPs): progress from a microRNA-based droplet digital PCR assay**

*Noemi Laprovitera<sup>1,7\*</sup>, Mattia Riefolo<sup>1\*</sup>, Elisa Porcellini<sup>1</sup>, Giorgio Durante<sup>1</sup>, Ingrid Garajova<sup>2</sup>, Francesco Vasuri<sup>3</sup>, Ariane Aigelsreiter<sup>4</sup>, Nadia Dandachi<sup>5</sup>, Giuseppe Benvenuto<sup>6</sup>, Federico Agostinis<sup>6</sup>, Silvia Sabbioni<sup>7</sup>, Ioana Berindan-Neagoe<sup>8</sup>, Chiara Romualdi<sup>6</sup>, Andrea Ardizzoni<sup>1</sup>, Davide Trerè<sup>1</sup>, Martin Pichler<sup>5</sup>, Antonia D'Errico<sup>1,3</sup> and Manuela Ferracin<sup>1</sup>*

<sup>1</sup> Department of Experimental, Diagnostic and Specialty Medicine (DIMES), University of Bologna, Bologna, Italy

<sup>2</sup> Medical Oncology Unit, University Hospital of Parma, Parma, Italy

<sup>3</sup> Pathology Unit, S. Orsola-Malpighi Hospital, Bologna, Italy

<sup>4</sup> Diagnostic and Research Institute of Pathology, Medical University of Graz, Graz, Austria

<sup>5</sup> Division of Oncology, Medical University of Graz, Graz, Austria

<sup>6</sup> Department of Biology, University of Padua, Padua, Italy

<sup>7</sup> Department of Life Sciences and Biotechnologies, University of Ferrara, 44121 Ferrara, Italy

<sup>8</sup> Research Center for Functional Genomics, Biomedicine and Translational Medicine, "Iuliu Hatieganu" University of Medicine and Pharmacy, 400337, Cluj-Napoca, Romania

\*equal contribution

**Running title:** CUP molecular site-of-origin prediction

**Corresponding author:** Manuela Ferracin, Department of Experimental, Diagnostic and Specialty Medicine (DIMES), University of Bologna, Via S. Giacomo, 14, Bologna, 40126, Italy, [manuela.ferracin@unibo.it](mailto:manuela.ferracin@unibo.it)

**Keywords:** MicroRNAs, metastasis, cancer of unknown primary, molecular diagnostics, droplet digital PCR

## Abstract

Metastasis is responsible for the majority of cancer-related deaths. Particularly challenging is the management of metastatic cancer of unknown primary site (CUP), whose tissue-of-origin (TOO) remains undetermined even after expensive investigations. CUP therapy is rather unspecific and poorly effective. Molecular approaches developed to identify CUPs' potential tissue-of-origin, can overcome some of these issues. In this study, we applied a pre-determined set of microRNAs (miRNAs) to infer the TOO of 53 metastatic cancers of unknown or uncertain origin. We designed a molecular assay to quantify 89 miRNAs at the copy number level, using EvaGreen-based Droplet Digital PCR. We assessed miRNA expression in 159 samples including primary tumors from 17 tumor classes (reference set), metastases of known and unknown origin. We applied two different statistical models for class prediction to obtain CUP's most probable TOOs. Specifically, we used the shrunken centroids using PAMR (Prediction Analysis of Microarrays for R) and the least absolute shrinkage and selection operator (LASSO) models. The molecular test was successfully applied to FFPE samples and provided a site-of-origin identification within one-week from the biopsy procedure. The most frequently predicted origins were gastrointestinal, pancreas, breast and lung. The assay was applied to multiple metastases from the same CUP, collected from different metastatic sites: the molecular prediction revealed an impressive agreement in site-of-origin prediction, intrinsically validating our assay. The final prediction was compared with the clinico-pathological hypothesis of primary site. Moreover, a panel of 14 miRNAs proved to have prognostic value and being associated with overall survival. Our study demonstrated that miRNA expression profiling in CUP samples could be employed as diagnostic and prognostic test. Our molecular analysis can be performed on-request, concomitantly with standard diagnostic workup and in association with genetic profiling, to offer valuable indication about the possible primary site, thereby supporting treatment decisions.

## Introduction

Cancer of Unknown Primary origin (CUP) describes newly diagnosed tumors presenting as metastatic cancers, whose primary site cannot be recognized after detailed standardized physical examinations, blood analyses, imaging and immunohistochemical (IHC) testing [1]. CUP biology represents a real riddle and several theories have been proposed to describe CUP origin. According to the two prevailing hypotheses, CUPs could either be metastatic tumors that originate from an undetectable, small, dormant or later regressed primary tumors or represent metastatic entities with no existing primary that derive from cancer stem cells and early acquire the metastatic ability [1].

Post-mortem investigations on CUP patients reported the identification of a primary tumor in about 75% of cases and highlighted the prevalent epithelial origin of CUPs. The most common primary sites were represented by lung, pancreas, hepatobiliary tract, kidney, colon, genital organs and stomach [2]. Population-based studies reported decreasing trends of CUP incidence in different countries in the last decade, possibly as a consequence of novel diagnostic techniques that improved primary site identification or a more consequent and widespread approach to follow standardized diagnostic work-up guidelines [3]. Nonetheless, incidence-rates still vary among different countries worldwide.

International guidelines for tumor treatment are essentially based on primary site indication. Therefore, CUP treatment requires a rather unspecific blind approach, which is very challenging for the treating physicians. As a consequence, CUPs are usually treated with empiric platinum-based chemotherapy regimens that are poorly effective. CUP patients have a short life-expectancy (average overall survival 4-9 months, 20% survive more than 1 year) that did not improve in the last decades. In the most recent CUP NCCN guidelines (v.2/2020), there are eleven different chemotherapy regimens indicated for adenocarcinoma and nine for squamous histology. However, these regimens remain empirical since they are mostly based on single-arm phase II clinical trials [4-6] and small randomized prospective trials [7-9]. In addition, the lack of primary tumor definition, prevent most patients to be treated in clinical practice with novel very effective treatment such as immunotherapy or molecular targeted therapies for which current registered indications are mostly disease-oriented. Finally, patients with occult primary tumors suffer a great psychological burden of an unidentified disease. The use of molecular tests that could identify the most probable site-of-origin or an approach based on personalized medicine, may be useful to assist in the selection of the best treatment options and potentially improve CUPs prognosis and survival.

The identification of druggable alterations in CUP tumors could improve the otherwise limited treatment options. Recently, several studies focused on the analysis of CUP mutational profiles [10-12]. A comprehensive retrospective analysis, using the 236-gene FoundationOne assay (Roche), explored the genomic profiles of 200 CUPs [11]. At least 1 clinically relevant genetic alteration was found in 96% of CUPs, with a mean of 4.2 alterations per tumor. Most frequently mutated genes were *TP53* (55%), *KRAS* (20%), *CDKN2A* (19%), *MYC* (12%), *ARID1A* (11%) and *MCL1* (10%). According to this study, potentially druggable mutations were discovered in 20% of CUPs. Varghese et al. identified the actionable mutations in a dataset of 150 CUPs analyzed with the MSK-IMPACT panel [12] and in another dataset of 200 CUPs from Ross et al. [11]. Potentially druggable alterations were present in 30% of CUP cases (FDA level 2-3 of evidence for actionability) [12]. Few patients in Varghese study were treated with targeted therapies, overall showing benefit and achieved longer survival.

Another way to improve the choice of CUP therapeutic options, is the prediction of CUP site-of-origin using molecular assays. This strategy is based on the observation that metastatic tumor cells retain some molecular characteristics of the tissue of origin, despite their going through de-differentiation and epithelial-mesenchymal transition programs. This tissue-specific molecular signature can be leveraged to obtain an inference about CUP origin. In the past decade, several molecular classifiers were developed. These classifiers were built based on gene expression profiles (GEP) [13-16], microRNAs [17, 18] or DNA methylation [19].

A number of studies reported evidences in favor of this hypothesis, showing a prolonged survival in patients treated with cancer-specific agents compared to standard chemotherapy [20, 21, 19, 22, 23]. Results from a prospective study on nearly 300 patients with unknown primary tumors who were treated according to GEP molecular prediction revealed a significant increase in median survival time (12.5 months) [24].

In addition, GEP proved a higher diagnostic accuracy compared to standard immunohistochemistry (IHC) staining in the identification of CUP primary site, especially in moderately or poorly differentiated cases [24, 25]. The most recent NCCN CUP Guidelines [26] support the use of gene expression profiling to get a diagnostic benefit in CUP management, though the achievement of a clinical benefit still needs to be determined. Results from the phase III clinical trial NCT03278600

could help to clarify the value of tissue-of-origin profiling in predicting primary site and directing therapy in CUP patients.

However, the analysis of GEP in archival formalin-fixed, paraffin-embedded (FFPE) tissues is limited by the quality of extracted RNA, which is usually low. Thus, the reported rate of technical success of GEP assays (i.e. CancerTypeID assay) is 85% [21]. On the contrary, microRNAs (miRNAs) are robustly detected irrespective of the quality of the tissue sample [27, 28] and are highly stable and resistant to RNAase degradation either in compromised archived clinical specimens [29, 30] or biological fluids [31]. Molecular miRNA profiling of FFPE samples could be successfully obtained from all the available samples [17, 32].

Independently from the molecular assay choice, assessing the true clinical benefit of molecular profiling is challenging because it relies on surrogate measures (correlation with IHC findings, clinical presentation or response to therapy), given that a real primary site identification is seldom available.

In a previous microarray-based study we identified a cancer-type specific miRNA signature able to predict metastatic tumor tissue-of-origin of CUPs among ten possible primary sites [17]. This predictive tool was employed in few occasions to provide clinicians with indications of a possible primary site [33]. However, microarray technology limitations prevent the execution of such analysis on a routine basis. To extend the analysis to more tumor types and overcome the technical limits of microarray technology, we developed a miRNA-based molecular assay for a rapid, on-demand molecular tumor characterization and primary site prediction [34]. Unlike previous assays, our test employs droplet digital PCR (ddPCR) technology to assess the absolute level of a pre-determined set of 89 miRNAs in FFPE tumor tissues. This assay is applied here to predict the most probable primary tissue(s) of a set of 53 cancers of unknown or uncertain origin, obtaining a broad spectrum of primary site predictions with different levels of confidence.

## **Methods**

### Patients and tumor samples

A total number of 159 formalin-fixed paraffin-embedded (FFPE) samples from 148 patients was collected for this study. Patients were diagnosed and treated at Sant'Orsola-Malpighi Bologna University Hospital, Italy (N=84), at the University Hospital of Ferrara, Italy (N=50) or at the Medical University of Graz, Austria (N=14). The study cohort consists of patients with tumors with a clearly recognized primary site (N=106) and patients with cancer of unknown or uncertain origin (CUPs,

N=53). A summary of samples and patients enrolled in the study is reported in **Table 1**. Primary tumors included samples obtained from the following tumor sites/types: lung (LUAD, adenocarcinoma, N=5 and LUSC, squamous cell carcinoma, N=3), pancreas (PAAD, exocrine adenocarcinoma, N=5), liver (LIHC, hepatocellular carcinoma, N=6), biliary tract (CHOL, cholangiocarcinoma, N=6), kidney (KICA, which includes kidney renal clear cell carcinoma or KIRC, N=5 and kidney renal papillary cell carcinoma or KIRP, N=3), colorectum (CRC, adenocarcinoma, N=7), testis (TGSC, germ cell seminomatous carcinoma, N=4), endometrium (UCEC, adenocarcinoma, N=5), stomach (STAD, adenocarcinoma, N=5), bladder (BLCA, transitional cell carcinoma, N=4), breast (BRCA, non-special type and lobular carcinoma, N=5), triple negative breast cancer (TNBC, N=3), prostate (PRAD, adenocarcinoma, N=5), melanoma (SKCM, melanoma of skin, N=7), head and neck (HNSC, squamous cell carcinoma, N=6) and gastrointestinal neuroendocrine carcinoma (GI-NET, N=5). We assessed 10 metastases of known origin, derived from lung, melanoma, stomach, prostate, head and neck, kidney, colon, breast, pancreas and endometrium. A total number of 53 CUPs were included in this study, specifically 43 retrospective and 10 prospective cases. Moreover, from 5 retrospective CUP patients we were able to obtain metastatic biopsies collected from multiple sites that were independently analyzed. CUP diagnosis was obtained after detailed clinical and pathological investigations. For each sample a full IHC panel was assessed at the time of diagnosis and the outcome was recorded. However, we need to underline that our collection of CUP samples is heterogeneous since it derives from patients that received the diagnosis in different time; specifically, 26% of them received the diagnosis of CUP between 2005-2009, 49% between 2010-2014 and 24% between 2015-2019.

For each sample, 10 µm thick tissue sections (N=2-5) were obtained. The first section was stained with hematoxylin-eosin (HE) and examined by an expert pathologist to select the tumor area, which was grossly dissected before RNA extraction. Tumor cell fraction was evaluated to select samples with at least 30% cellularity. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee Center Emilia-Romagna Region – Italy (protocol 130/2016/U/Tess) and Medical University of Graz (vote no. 30-520 ex 17/18). Prospective patients provided written informed consent. Detailed pathological characteristics of cancer patients are available in **Supplementary Table 1** and **Table 2**.

#### RNA extraction and cDNA conversion

Total RNA, including microRNAs, was isolated from the tumor FFPE sections using miRNeasy FFPE kit (Qiagen, Cat No. 217504; miRNeasy FFPE Handbook Qiagen, HB-0374-005). Deparaffinization was performed with xylene as described in *Appendix A*. Further instructions were followed from step 11 of the *Protocol: Purification of Total RNA, Including miRNA, from FFPE Tissue Sections*. RNA was eluted in 20-30  $\mu$ L and frozen at  $-80^{\circ}\text{C}$ . RNA yield and quality were assessed with NanoGenius Spectrophotometer (ONDA Spectrophotometer). All samples were suitable for the molecular testing.

RNA conversion to cDNA was performed using the miRCURY LNA RT Kit (Qiagen, Cat No. 339340; miRCURY LNA miRNA PCR Handbook, HB-2431-002). The 10  $\mu$ L-reaction mix was prepared for each sample mixing: 2  $\mu$ L of 5x Reaction Buffer, 4.5  $\mu$ L of nuclease free water, 1  $\mu$ L of enzyme mix, 0.5  $\mu$ L of UniSp6 RNA spike-in and 2  $\mu$ L of diluted RNA (10 ng of total RNA). The resulting cDNA was stored in LoBind DNA Eppendorf tubes (Eppendorf, 0030108051) at  $-20^{\circ}\text{C}$ . For each sample, a RT-qPCR was performed as quality control step using miRCURY LNA miRNA PCR Assays (Qiagen) to test UniSp6 (Cat No. YP00203954) and SNORD44 (Cat No. YP00203902) targets. UniSp6 threshold cycle (Ct) informs about the RT reaction efficiency. SNORD44 was tested to assess RNA integrity and amplifiability and to establish the cDNA dilution prior to digital droplet PCR (ddPCR) analysis. For SNORD44 Ct ranging 24-30 (threshold set at 160), cDNA was diluted 1:50; for Ct below 24, cDNA was diluted 1:100-1:200 and when Ct was higher than 30, the RT was repeated again using undiluted RNA and qPCR analysis repeated. cDNA was further diluted 1:10 in miR-21-5p and UniSP6 wells. Applying these criteria, we prevented ddPCR saturation problems or low miRNA expression levels in ddPCR analysis.

#### Droplet Digital PCR and data analysis

Pre-spotted custom plates (96-well format) were designed to comprehend 89 different miRCURY LNA miRNA primers (Qiagen), three assays for small nuclear or nucleolar RNAs as reference candidates (SNORD44, SNORD48 and snRNAU6), two interplate calibrator assays (UniSp3), a control plate assay (UniSP6) and a no template control (NTC) as described in [34] (miRNA list in **Supplementary Table 2**).

EvaGreen-based Droplet digital PCR was performed as described in [35, 36, 34]. Thermal cycling conditions were:  $95^{\circ}\text{C}$  for 5 min, then 40 cycles of  $95^{\circ}\text{C}$  for 30 s and  $58^{\circ}\text{C}$  for 1 min (ramping rate reduced to 2%), and three final steps at  $4^{\circ}\text{C}$  for 5 min,  $90^{\circ}\text{C}$  for 5 min and a  $4^{\circ}\text{C}$  infinite hold. Droplet selection was performed individually for each well using QuantaSoft software v 1.7 (Bio-Rad). Final

miRNA amounts (copies/ul) were obtained and normalized on 50<sup>th</sup> percentile expression using GX v.14.9.1 software (Agilent Technologies).

### Tissue-of-origin prediction

Primary tumors (N=96) were used as training set as previously described [17]. Nearest shrunken centroids (NSC) algorithm [37] using the Prediction Analysis of Microarray for R (PAMR) tool [37] and the Least Absolute Shrinkage and Selection Operator (LASSO) model [38] were used to build up the classifiers. The PAM threshold was set to 0 leading to a classifier based on 88 miRNAs while the LASSO threshold was set to 0.019 leading to a classifier based on 48 miRNAs. Then, these classifiers were used to predict known and unknown/uncertain metastases tissue-of-origin. Both predictive models assign to every metastatic tumor a probability to be originated from each primary site. The variable gender was also taken into account to exclude not compatible molecular predictions (TGSC/PRAD in females and OV/UCEC in males). Results were compared with the indications of a possible primary site suggested by standard diagnostic workup and clinico-pathological assessment. Bootstrap approach (with N=100) was used to assess the performance (error rate) of the models in the training set.

### Cluster analysis

The hierarchical cluster analysis (with complete-linkage rule and Pearson Correlation distance) of primary tumor samples was obtained using GeneSpring GX v.14.9.1 software (Agilent Technologies) on ddPCR data normalized on the 50<sup>th</sup> percentile.

### Survival analysis

Univariate survival analysis was performed using Kaplan-Meier curves and the log rank test as implemented in survMisc R package. Overall survival (OS) was calculated considering the time lagging between diagnosis and death for any cause or the last follow-up. For each miRNA the optimal cut-off was estimated as the threshold on the ROC curves that maximize the sum of specificity and sensitivity in predicting CUP patients. Results were reported as *p* value, hazard ratio (HR) and 95% confidence intervals (CI). A *p* value  $\leq 0.05$  was considered significant.

## **Results**

### ***Multi-miRNA testing on archive samples with droplet digital PCR***



We implemented a miRNA-signature for tumor primary site prediction integrating two published signatures [17, 39] plus 10 additional miRNAs (miR-661, miR-24-3p, miR-21-5p, miR-16-5p, miR-320a, miR-224-5p, miR-423-5p, miR-25-3p, miR-331-3p and miR-103a-3p) as detailed in **Supplementary Table 2**.

FFPE tissue is the most commonly available source of tumor material for molecular profiling in the clinical setting and miRNAs are extremely stable in FFPE blocks. Therefore, we developed an on-demand multi-miRNA assay capable of testing the absolute levels of 89 miRNAs in a 2-days timeframe compatible with standard diagnostic workup and with the amount of available material. The multi-miRNA assay is based on absolute miRNA quantification with EvaGreen Dye Droplet Digital PCR technology [34]. From a technical point of view, the assay provided good quality results for all tested archive FFPE samples. RNA was extracted from 2-5 slices of tumor FFPE blocks, then the tumor area was identified by experienced pathologists and macrodissected. An amount of 10 ng is sufficient to test all miRNAs in a single experiment, thus confirming the feasibility in a diagnostic setting.

We obtained the absolute copy number for all miRNAs included in our panel in the same Droplet Digital PCR experiment, with identical experimental conditions (annealing temperature and amount of primers), only adjusting the amount of input cDNA for miR-21-5p and UniSP6.

With the aim of establishing a reference set for cancer of unknown origin molecular profiling, we tested 96 primary tumors with our multi-miRNA assay, comprising 16 different tumor types and 19 histological classes, focusing on the most common CUP's sites of origin identified at autopsy [40]. We obtained the expression matrix of the primary tumor dataset, constituted by tumors belonging to 19 different classes: LUAD, LUSC, PAAD, LIHC, CHOL, KIRC, KIRP, STAD, CRC, TGSC, OV, UCEC, BLCA, BRCA, TNBC, PRAD, SKCM, GI-NET and HNSC. An overview of the primary tumor samples for each histological subtype included in this study is reported in **Table 1**.

### ***Analysis of miRNA expression patterns***

We evaluated the average levels of normalized expression of the 89-miRNA signature in the nineteen primary tumor types with cluster analysis (**Figure 1**). Each tumor type displays a peculiar pattern of miRNA expression, as expected. Nonetheless, we found some unexpected similarities and divergences among tumor types, which are worth mentioning. Specifically, miRNA expression of STAD and CRC was found to be consistently overlapping and partially intermixed with other gastrointestinal tumors (PAAD and GI-NET), as reported also in previous reports [17, 39, 41]. Due to

this miRNA expression similarity, we decided to consider them as a single class (STAD-CRC) for molecular prediction. Similarly, kidney renal clear cell (KIRC) and papillary cell carcinomas (KIRP), showing similar miRNA expression patterns, were combined in the tumor class KICA. Tumors in female reproductive-system organs (OV and UCEC) were found to express similar yet distinct miRNA patterns as previously observed [17, 39, 41]. Moreover, lung cancers (both LUAD and LUSC) share a portion of their signatures with TNBC but not with other breast cancer subtypes (ER+, PR+, HER2+ tumors). TNBC shows a largely different pattern of miRNA expression when compared to other breast cancers, showing an unexpected similarity with HNSC instead. We could speculate that a common etiology associated to human papilloma virus (HPV) infection has been reported in both these tumor types [42-44]. Overall, this signature confirmed its potential in discriminating among 17 different tumor classes.

### ***CUP predictive model generation***

The final primary site prediction was performed using 88 out of 89 miRNA assays of our panel as miR-122-5p was excluded from the classifier due to its strong signal generated by the liver microenvironment in metastatic samples (**Supplementary Figure 1**).

We applied the Nearest Shrunken Centroids (NSC) using PAMR [37] and the Least Absolute Shrinkage and Selection Operator (LASSO) predictive models [38] developed by Tibshirani's lab to our training set of primaries. To assess the performance of the predictive models on the training set we used a bootstrap approach. Error rates for each tumor class for both models are reported in **Supplementary Table 3**. Notably, the overall error rates for PAMR and LASSO were 34.3% and 35.1%, respectively. However, 11 of the 17 tumor classes (LIHC, LUSC, HNSC, BRCA, OV, CHOL, KICA, GI-NET, TGSC, STAD-CRC, SKCM, LUAD, UCEC, PRAD) had error rates significantly lower for both models (17.3% for PAMR and 22.7% for LASSO). Of note, PAMR seem to be considerably more accurate in the prediction of BRCA compared to LASSO, with no error reported; on the contrary, LASSO seem to be slightly more precise in the identification of UCEC, LUAD and SKCM. Both models had higher error rates in identifying correctly BLCA, PAAD, TNBC, HNSC, OV and CHOL classes; this might be explained by the low specificity of miRNA signatures for these primaries or by their small sample size. From these results it is clear that the two models behave similarly on some classes and complementarily in some others, thus we decided to take advantage of both classifiers for the second step of molecular prediction.

A small set of metastases of known origin (N=10) was assessed for molecular prediction (test set **Table 1**). Considering the two top predicted classes we obtained an accuracy of 80% for PAMR and of 60% for LASSO, as reported in **Supplementary Table 4**.

### ***CUP primary site prediction***

Finally, both models were used to predict the primary site of 53 cancers of unknown/uncertain origin (CUPs). Given the tumor frequency, this is a remarkably large collection of cancers of unknown primary site whose histopathological and immunohistochemistry characteristics are detailed in **Supplementary Table 1** and **Table 2**. The prediction outcome is represented in **Figure 2** in which the top two primary sites predicted by both models for each CUP sample are reported. Using PAMR, the molecular prediction of 41 out of 53 CUPs (77.3%) predicted a primary site with a probability higher than 60%, half of them even higher than 90%. Using LASSO, the molecular prediction of 13 out of 53 CUPs (24.5%) predicted the first primary site with a probability higher than 60%, and 3 higher than 90%. We identified a subgroup of CUP patients for which it was very challenging to point out a tissue-of-origin using both models, thus suggesting an exceptionally undifferentiated phenotype. Of note, in the majority of cases LASSO and PAMR predictions were in agreement. Final molecular predictions are reported in **Table 2**. The most probable primary site which were selected based in the agreement between the two models and compatibility with the clinical/pathological information available. Considering the final predicted sites reported in **Table 2**, the most common tissues-of-origin were STAD-CRC (24.6%), PAAD (24.5%), BRCA (18.9%), LUAD or LUSC (15.1%), CHOL (15.1%), TNBC (8.2%) and HNSC (7.5%) and others at lower rates. Of note, no CUP was predicted to originate from TGSC or PRAD.

Interestingly, from five CUP patients we obtained a number of samples (N=2-4) derived from multiple metastatic sites, which were all tested with our assay. These samples were used to evaluate the consistency of our prediction. Symbolic is the case of a patient (#B) with an initial diagnosis of cancer of unknown primary (later attributed to a breast origin) from which we obtained a total number of four samples (CB098, CB100, CB101, CB102). In particular, CB098 and CB100 were obtained from two lymph nodes resected in 2010, while CB101 and CB102 derived from an invasive ductal breast cancer identified two years later, which was recognized as the primary site. Both PAMR and LASSO agreed to predict it as a BRCA or TNBC (**Table 2**). However, CB100 was predicted as LUAD (first) or TNBC (second) by PAMR classifier (**Figure 2**), probably due to the lower tumor cell fraction in this sample and the reported similarity in miRNA expression between breast and lung

cancers [17]. Molecular predictions for the multiple metastases of the other patients (#E, #F, #Q and #R) reported concordant results for both models, in agreement with clinical or pathological hypotheses. Specifically, for #E (CB105 and CB106) and #F (CB108 and CB109) both models agreed to predict a gastrointestinal origin (STAD-CRC), which was also the first clinicopathological hypothesis. CB108 from #F patient had a different indication as most probable tissue-of-origin with PAMR classifier (breast); however, being derived from the bone it is probable that the sample had a compromised integrity. Molecular prediction for #R (CB121 and CB122) pointed out to a biliopancreatic origin, while for #Q (CB119 and CB120) the two metastatic samples were predicted to have the same origin, which was in this case lung or head and neck. CUP prediction probabilities with PAMR and LASSO models are reported in **Supplementary Table 5**.

### ***Association of microRNAs with CUP patients' overall survival***

We tested the performance of our 89-miRNA panel as prognostic test for CUP patients. Survival information was available for 34 CUP patients included in this study. We performed a survival analysis to test the association of miRNA expression with overall survival (**Supplementary Table 6**) finding 14 miRNAs with significant prognostic effect on CUP patients' OS (**Table 3 and Figure 3**). The association between survival probability and miRNA expression was negative for 6 miRNAs ( $HR > 1$ ) and positive for 8 miRNAs ( $HR < 1$ ).

In particular, the miRNAs whose higher expression is associated with worse prognosis are miR-21-5p ( $p= 0.001$ ), miR-375 ( $p=0.03$ ), miR-27b-3p ( $p=0.03$ ), miR-96-5p ( $p= 0.04$ ), miR-423-5p ( $p=0.04$ ) and miR-214-3p ( $p= 0.05$ ). On the contrary, 8 miRNAs are positively associated with a prolonged survival: miR-124-3p ( $p= 0.0002$ ), miR-9-3p ( $p= 0.01$ ), miR-149-5p ( $p= 0.01$ ), miR-372-3p ( $p= 0.03$ ), miR-485-5p ( $p= 0.03$ ), miR-25-3p ( $p= 0.03$ ), miR-181a-2-3p ( $p= 0.03$ ) and miR-10b-5p ( $p= 0.04$ ).

## **Discussion**

The identification of the tissue-of-origin in metastatic cancers strongly relies on clinical information and histology as well as immunohistochemical evaluations but this diagnostic workup is sometimes ineffective and a fraction of primaries remains unidentified. Epitome of this scenario is metastatic cancer of unknown primary site (CUP), which presents by definition as an advanced cancer whose site of origin is not detectable nor presumable, despite an intensive clinical and pathological diagnostic workup [1]. CUPs represent an enigma both at the biological and pathological level, and an important under-researched clinical problem.

In the past decade, several molecular tests based on gene expression (GEP), microRNA or DNA methylation profile, were developed to improve primary site identification in cancers of unknown/uncertain origin. The underlying premise for these molecular profiling assays (reviewed in [45] and *Laprovitera et al. (Cancers, accepted)*) is that metastatic tumors preserve specific molecular signatures that match their primary site and can be used to identify their site-of-origin. Overall, these methods reach a prediction accuracy that ranges from 80 to 95% and have the potential to improve the diagnostic workup of CUP patients and guarantee the access to more therapeutic options. Indeed, NCCN occult primary guidelines recently assessed CUP molecular profiling as a potential provider of clinical benefit for patients. At the present time, CUP molecular profiling's clinical utility need to be determined on a case-by-case basis, and clinical validation in large randomized phase III trials is still missing.

In this study, we developed a molecular assay to assess the expression of 89 miRNAs in tumor FFPE samples by using Droplet Digital PCR (ddPCR) and infer CUP primary site [34]. Our miRNA panel was determined merging two cancer-specific miRNA signatures previously identified in two microarray-based studies [17, 39]. To prevent the costs of large-scale technologies such as microarrays or sequencing, we opted for a focused number of selected miRNAs and the use of ddPCR technology. This assay allows the on-demand quantification of a focused panel of miRNAs per sample, at an affordable cost and in a 2-days timeframe. Droplet digital PCR technology provides miRNA absolute quantification without the requirement of standard curves, efficiency correction approaches or technical replicates typical of traditional quantitative PCR approaches [46]. In addition, EvaGreen-based ddPCR allows to precisely detect target miRNAs at levels down to 1 copy/ $\mu$ L [47].

As we hypothesized, an approach based on miRNA expression instead of gene expression profiles is very convenient since we were able to successfully analyze the totality of FFPE samples in our cohort (100% success rate), with no excluded sample due to technical issues.

In this study we analyzed the 89-miRNA profiles of 159 FFPE samples, including 53 CUPs, and successfully obtained a primary site prediction for all patients. We obtained a good prediction accuracy rate in metastatic cancers of known origin, and highly consistent results when assessing multiple metastases derived from the same CUP patient. These two settings provided an intrinsic validation of our combined predictive models.

As for CUP predictions, we observed consistency between our prediction outcomes and clinical and histo-pathological hypotheses, when they were available. In addition, we were able to successful analyze all 159 FFPE samples, with no excluded sample due to technical issues. The employment of

two predictive models allowed us to obtain stronger results when both systems pointed out to the same tissue-of-origin. Of note, some CUPs were molecularly predicted as LUAD with a negativity for TTF1, which defines a subgroup of LUAD with unfavorable outcomes [48]. Our results provide further evidence of the translational potential of CUP molecular testing in general and miRNA testing in particular. With no intention to replace IHC testing, molecular assays can support the pathologists in narrowing the spectrum of possible primary sites of undifferentiated metastatic tumors. When no pathological hypothesis can be formulated, the miRNA-based molecular assay could aid the oncologists in their therapeutic choice, despite being necessary to demonstrate a benefit in a clinical setting.

The droplet digital PCR, miRNA-based assay herein applied has an accuracy comparable with other commercialized molecular profiling assays, but overcomes some limits of previous tools. Our molecular classifiers have the advantage to cover a wide variety of primary cancers, among the most likely to be CUP's sites-of-origin; in particular, we can discriminate between 17 primary tumor subtypes. The ability to cover such number of tumor classes is an advantage if compared to other commercialized molecular assays e.g. the 10-genes qPCR assay (Veridex) that can classify only six different tumor types. Our prediction outcome on CUPs mostly overlap the frequency rates identified in post-mortem autopsy studies: lung (27%), pancreas (24%), liver or bile duct system (8%), kidney or adrenal (8%) or colon (7%) [40].

Three molecular assays were recently approved for CUP diagnostics in US: Pathwork Tissue of Origin Test (Pathwork Diagnostics), Cancer-TYPE ID (bioTheranostics), and miRview mets<sup>2</sup> (Rosetta Genomics). The first is a microarray-based system that assessing the gene expression profiles (GEP) of 2000 genes claim to distinguish up to 15 tumor types. CancerTYPE ID is another GEP-based assay which evaluates by RT-qPCR the expression of a 92-gene signature and identifies the primary origin of up to 30 tumor types. Finally, miRview mets<sup>2</sup> system, assessing the expression of 64 miRNAs by RT-qPCR, is able to distinguish up to 26 tumor types.

However, these assays included primary tumors that have little or no connection with CUPs. Our molecular tool is able to cover a high number of tumor classes, selected as the most common CUP tissues of origin. Our assay has a 100% success rate and requires a 2-days working time, which is compatible with a standard diagnostic workup and consistently shorter compared to other commercial assays that present a turnaround time of 5-11 days.

In addition to being faster, targeted and cost-effective in primary site identification; our assay could be easily combined with the analysis of druggable alterations, to select CUP therapy. However, further prospective clinical studies are necessary to evaluate their use in the clinics and to demonstrate its possible impact on CUP patients' survival.

In conclusion, our study demonstrated that digital miRNA expression profiling of CUP samples has the potential to be employed in a clinical setting in FFPE tissue. Our molecular analysis can be performed on-request, concomitantly with the standard diagnostic workup and in association with genetic profiling, to offer valuable indication about the possible primary site thereby supporting treatment decisions.

## **Declarations**

### **Ethics approval and consent to participate**

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee Center Emilia-Romagna Region – Italy (protocol 130/2016/U/Tess) and Medical University of Graz (vote no. 30-520 ex 17/18). Prospective patients provided written informed consent.

### **Consent for publication**

All authors of the manuscript have read and agreed to its content.

### **Availability of data and material**

Droplet digital PCR data are available upon request.

### **Competing interests**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### **Funding**

The research leading to these results has received funding from Fondazione Italiana per la Ricerca sul Cancro (AIRC) under IG 2016 - ID. 18464 project – P.I. Ferracin Manuela.

### **Authors' contributions**

Conception and design: M. Ferracin

Development of methodology: N. Laprovitera, M. Riefolo, E. Porcellini, M. Ferracin, C. Romualdi

Acquisition of data: I. Garajova, F. Vasuri, A. Aigelsreiter, N. Dandachi, I. Berindan-Neagoe, A. Ardizzoni, D. Trerè, A. D'Errico

Analysis and interpretation of data: N. Laprovitera, M. Riefolo, E. Porcellini, F. Agostinis, G. Durante, G. Benvenuto, A. D'Errico, M. Pichler, M. Ferracin, C. Romualdi

Writing, review, and/or revision of the manuscript: N. Laprovitera, M. Riefolo, M. Pichler, A. D'Errico, M. Ferracin

Study supervision: M. Ferracin

### **Acknowledgments**

MR was a fellow of Fondazione Famiglia Parmiani, Bologna (Italy). NL is supported by eDIMES Lab funds from Bologna University.



## Figure legends

**Figure 1 – Cluster analysis of primary tumors.** Heatmap representing 89-microRNA expression in nineteen different classes of primary carcinomas. Averaged, normalized miRNA levels in each tumor type were used for clustering analysis. Green indicates low expression, red indicates high expression.

**Figure 2 – Prediction outcome of cancers of unknown origin using PAMR NSC and LASSO classifiers.** For each of the 53 CUP sample (on the y-axis) the two top predicted primary tumors (x axis) are highlighted. PAMR first and second molecular predictions are reported with dark and light-blue squares, respectively. LASSO first and second molecular predictions are reported in dark- and light-orange, respectively. A diamond in the cell indicates those tissues-of-origin that are consistent with pathological and/or clinical information.

**Figure 3 - Kaplan–Meier OS curves based on the expression of 14 miRNAs in CUP patients.** Survival plots showing significantly different OS curves in high- and low-miRNA expression cases. The threshold for each miRNA was established based on the best performing value at ROC analysis. A higher expression of 6 miRNAs is associated with shorter CUP survival and of 8 miRNAs with prolonged survival.

### Supplementary information

**Supplementary Figure 1 - Plot of miR-122-5p expression in primary and metastatic tumors.**

Normalized miR-122-5p expression was evaluated in liver and bile duct primary tumors, known to express this miRNA at high levels, and in metastatic tumors of known/unknown origin whose biopsy was obtained from the liver tissue or other sites. Liver metastases of known/unknown origin show high levels of miR-122-5p if compared to those derived from other sites, which is due to the very abundant expression of miR-122 in liver cells and its release in the tumor microenvironment.

**Supplementary Table 1 – Clinic-pathological features of 159 samples.**

**Supplementary Table 2 – List of miRNA assays in the custom ddPCR plate.**

**Supplementary Table 3. Error rates of the PAMR and LASSO models for each tumor class.**

**Supplementary Table 4. Primary site prediction in metastases of known origin.**

**Supplementary Table 5. CUP probabilities with PAMR and LASSO classifier models**

**Supplementary Table 6. Association of miRNA expression with CUP overall survival (all miRNAs).**

**Table 1.** Summary of samples and patients enrolled in the study

|                               |             |    | Primaries |    | Metastases |    | CUPs  |       |
|-------------------------------|-------------|----|-----------|----|------------|----|-------|-------|
| Characteristics               |             |    | n         | %  | n          | %  | n     | %     |
| <b>Patients</b>               |             |    | n = 148   |    | 92         |    | 10    | 46    |
| <b>Prospective</b>            |             |    |           |    |            |    | 10    | 22    |
| <b>Retrospective</b>          |             |    |           |    | 92         |    | 10    | 43 93 |
| <b>Samples</b>                |             |    | n = 159   |    | 96         |    | 10    | 53    |
| <b>Sex</b>                    |             |    |           |    |            |    |       |       |
|                               | Male        | 78 | 48        | 52 | 3          | 30 | 26    | 57    |
|                               | Female      | 64 | 32        | 35 | 5          | 50 | 27    | 59    |
|                               | ND          | 17 | 16        | 17 | 2          | 20 | 0     | 0     |
| <b>Age, years</b>             |             |    |           |    |            |    |       |       |
|                               | Median      |    | 66        |    | 71         |    | 67    |       |
|                               | Range       |    | 44-85     |    | 60-86      |    | 42-87 |       |
|                               | ND          |    | 62        |    | 4          |    | 0     |       |
| <b>Primary tumor classes:</b> |             |    |           |    |            |    |       |       |
|                               | BLCA        |    | 4         |    |            |    |       |       |
|                               | BRCA        |    | 5         |    |            |    |       |       |
|                               | CHOL        |    | 6         |    |            |    |       |       |
|                               | CRC         |    | 7         |    |            |    |       |       |
|                               | GI-NET      |    | 5         |    |            |    |       |       |
|                               | HNSC        |    | 6         |    |            |    |       |       |
|                               | KIRC        |    | 5         |    |            |    |       |       |
|                               | KIRP        |    | 3         |    |            |    |       |       |
|                               | LIHC        |    | 6         |    |            |    |       |       |
|                               | LUAD        |    | 6         |    |            |    |       |       |
|                               | LUSC        |    | 3         |    |            |    |       |       |
|                               | OV          |    | 6         |    |            |    |       |       |
|                               | PAAD        |    | 5         |    |            |    |       |       |
|                               | PRAD        |    | 5         |    |            |    |       |       |
|                               | SKCM        |    | 7         |    |            |    |       |       |
|                               | STAD        |    | 5         |    |            |    |       |       |
|                               | TGSC        |    | 4         |    |            |    |       |       |
|                               | TNBC        |    | 3         |    |            |    |       |       |
|                               | UCEC        |    | 5         |    |            |    |       |       |
| <b>Metastatic sites</b>       |             |    |           |    |            |    |       |       |
|                               | bone        |    |           |    |            |    | 2     |       |
|                               | bone marrow |    |           |    |            |    | 1     |       |
|                               | brain       |    |           |    | 1          |    | 2     |       |

|              |   |    |
|--------------|---|----|
| breast       |   | 3  |
| cerebellum   |   | 1  |
| colon        | 1 | 1  |
| dermis       |   | 1  |
| duodeno      |   | 1  |
| kidney       |   | 1  |
| liver        | 4 | 12 |
| lung         | 1 | 2  |
| lymph node   |   | 14 |
| muscle       |   | 1  |
| ND           |   | 1  |
| ovary        |   | 1  |
| pericardium  | 1 |    |
| pleura       |   | 5  |
| prostate     |   | 2  |
| skin         | 1 |    |
| soft tissues |   | 2  |
| stomach      | 1 |    |
| thyroid      |   | 1  |

---

**Table 2. Prediction outcome in cancer of unknown primary site**

| Sample ID | Sex | Multiple metastases | Age at diagnosis | Tumor cellularity (%) | Status        | Biopsy site | Histotype               | K WS    | K7      | K20 | Other IHC testing  | Pathological hypothesis | Clinical hypothesis | Molecular prediction (most probable primary tumor) |
|-----------|-----|---------------------|------------------|-----------------------|---------------|-------------|-------------------------|---------|---------|-----|--|-------------------------|---------------------|--|
| CB002     | F   |                     | 65               | 50                    | retrospective | liver       | adenocarcinoma          | ND      | POS     | NEG | CA125+, chromogranin-, ER-, GATA3-   | ND                      | ND                  | STAD-CRC / BRCA                                    |
| CB003     | F   |                     | 81               | 60                    | retrospective | lymph node  | carcinoma               | POS     | ND      | ND  | S100-, CD10-, TTF1-, ER-, PR-, HMB45-, GATA3-  | ND                      | ND                  | HNSC / LUSC  |
| CB011     | F   |                     | 75               | 85                    | retrospective | lymph node  | carcinoma               | POS     | POS     | ND  | WT1+, vimentin+, chromogranin +/-, synaptophysin +/-, calretinin-, CD10-, ER-, PR-, p63-, CD45-, CDX2- | mullerian or kidney     | ND                  | PAAD   |
| CB012     | M   |                     | 76               | 30                    | retrospective | bone marrow | adenocarcinoma          | POS/NEG | NEG     | NEG | S100+, PSA +/-, CK14-, CK18-, CK 19-, ER-, PR-, HMB45-, MUC1+, TTF1-, vimentin-                        | prostate                | melanoma            | CHOL / LIHC  |
| CB013     | F   |                     | 71               | 70                    | retrospective | lymph node  | mucinous adenocarcinoma | POS     | NEG/POS | NEG | ER-, PR-, TTF1-, CDX2-   | gastrointestinal        | ND                  | STAD-CRC / PAAD                                    |

|              |   |  |    |    |               |            |                          |                |                |            |  |                      |        |      |
|--------------|---|--|----|----|---------------|------------|--------------------------|----------------|----------------|------------|--|----------------------|--------|------|
| <b>CB014</b> | F |  | 47 | 50 | retrospective | lymph node | papillary adenocarcinoma | <b>POS/NEG</b> | <b>POS/NEG</b> | <b>NEG</b> | vimentin +/-, actin-, CK14+, ER-, PR-, HER2-, TTF1-, thyroglobulin-, WT1-, GATA3+, P40-, PAX8-         | TN breast or thyroid | ND     | TNBC |
| <b>CB033</b> | F |  | 77 | 60 | prospective   | lymph node | carcinoma                | <b>POS</b>     | <b>POS</b>     | <b>NEG</b> | CK 5-6+, CK14+, GATA3+/-, ER-, PR-, AR, HER2-, P63+, P40+, synaptophysin-, PAX8-, napsin A-WT1-, TTF1- | sudoriparous gland   | breast | BRCA |
| <b>CB053</b> | F |  | 60 | 75 | prospective   | liver      | adenocarcinoma           | ND             | <b>POS</b>     | <b>NEG</b> | TTF1+, ALK-, KRAS G34T (pyrosequencing)  | lung                 | ND     | BRCA |
| <b>CB054</b> | F |  | 59 | 70 | prospective   | lung       | carcinoma                | <b>POS</b>     | ND             | ND         | GATA3+, ER-, PR-, HER2-, Ki67 96%, p40-, p63-, TTF1-, CDX2-  | breast               | ND     | BRCA |
| <b>CB055</b> | F |  | 49 | 50 | prospective   | lymph node | adenocarcinoma           | <b>POS</b>     | <b>POS</b>     | <b>POS</b> | CDX2+, chromogranin-, synaptophysin-, CD56-, HER2-, MSI-   | gastrointestinal     | ND     | CHOL |
| <b>CB061</b> | M |  | 60 | 70 | retrospective | kidney     | adenocarcinoma           | <b>POS</b>     | <b>POS</b>     | ND         | AR+, CD10-, OCT4-, PSA-, RCC-, TTF1-, GATA3-, NKX34.1-   | breast or kidney     | ND     | BRCA |

|              |   |  |    |    |               |                 |                    |            |            |            |   |                      |       |          |
|--------------|---|--|----|----|---------------|-----------------|--------------------|------------|------------|------------|---|----------------------|-------|----------|
| <b>CB062</b> | M |  | 87 | 80 | prospective   | seminal vesicle | carcinoma          | <b>POS</b> | <b>NEG</b> | <b>NEG</b> | CDX2+, SMA-, CD34-, desmin-, HER2-, Ki67:50%, MART1-, MSI-, NKX3.1-, S100-  | gastrointestinal     | ND    | CHOL     |
| <b>CB064</b> | M |  | 58 | 65 | prospective   | liver           | squamous carcinoma | ND         | <b>NEG</b> | <b>NEG</b> | P63+, CK14+, chromogranin -, synaptophysin -, CD45-, S100-  | ND                   | ND    | HNSC     |
| <b>CB071</b> | F |  | 64 | 60 | prospective   | liver           | carcinoma          | ND         | <b>POS</b> | <b>NEG</b> | CEA +/-, GATA3-, ER-, PR-/+, TTF1-, PAX8-   | biliar duct          | ND    | CHOL     |
| <b>CB090</b> | F |  | 63 | 30 | prospective   | duodeno         | adenocarcinoma     | ND         | <b>POS</b> | <b>NEG</b> | CDX2 +/-, MUC 1+, MUC 2-, MUC 5AC+, MUC 6+/-, HER2-, MSI-, synaptophysin -, TTF1-                                   | gastrointestinal     | ND    | STAD-CRC |
| <b>CB095</b> | F |  | 70 | 80 | prospective   | soft tissue     | squamous carcinoma | ND         | <b>POS</b> | <b>NEG</b> | CA125-, CA15.3-, CDX2 +/-, desmin-, ER-, GATA3-, MUC1+, P40-, PAX-8-, PD-L1 70%, ER-, S100-, synaptophysin -, TTF1- | ND                   | ND    | BLCA     |
| <b>CB097</b> | F |  | 68 | 90 | retrospective | liver           | neuroendocrine     | ND         | ND         | ND         | chromogranin +, synaptophysin +, CDX2-, gastrin-,   | gastrointestinal NET | liver | PAAD     |

|              |   |     |    |    |                   |               |                            |            |            |            |   |        |        |                |
|--------------|---|-----|----|----|-------------------|---------------|----------------------------|------------|------------|------------|---|--------|--------|----------------|
|              |   |     |    |    |                   |               |                            |            |            |            | glucagon-,<br>insulin-, TTF1-   |        |        |                |
| <b>CB098</b> | F | B01 | 42 | 80 | retrospecti<br>ve | lymph<br>node | squamous cell<br>carcinoma | <b>POS</b> | <b>POS</b> | <b>POS</b> | CD10-, CEA+,<br>GATA3-,<br>HER2-, Ki67<br>90%,<br>Mammaglobin<br>-, ER-, PR-, AR-<br>, HER2-, BRAF<br>V600E-                              | breast | breast | TNBC           |
| <b>CB100</b> | F | B02 | 42 | 40 | retrospecti<br>ve | lymph<br>node | squamous cell<br>carcinoma | <b>POS</b> | <b>POS</b> | <b>POS</b> | CD10-, CEA+,<br>GATA3-,<br>HER2-, Ki67<br>90%,<br>Mammaglobin<br>-, TTF1-, ER-,<br>PR-, AR-,<br>HER2-, BRAF<br>V600E-, EBV-<br>(ISH)      | breast | breast | TNBC /<br>BRCA |
| <b>CB101</b> | F | B03 | 42 | 80 | retrospecti<br>ve | breast        | adenocarcinom<br>a         | <b>POS</b> | <b>POS</b> | <b>POS</b> | CD10-, CEA+,<br>GATA3-,<br>HER2-, Ki67<br>70%,<br>Mammaglobin<br>-, CatepsinK-,<br>ER-, PR-, AR-,<br>HER2-, BRAF<br>V600E-, EBV-<br>(ISH) | breast | breast | TNBC /<br>BRCA |
| <b>CB102</b> | F | B04 | 42 | 85 | retrospecti<br>ve | breast        | adenocarcinom<br>a         | <b>POS</b> | <b>POS</b> | <b>POS</b> | CD10-, CEA+,<br>GATA3-,<br>HER2-, Ki67<br>60%,<br>Mammaglobin<br>-, CatepsinK-,   | breast | breast | TNBC /<br>BRCA |



|              |   |     |    |    |               |                      |                            |            |            |            |   |                              |                              |          |
|--------------|---|-----|----|----|---------------|----------------------|----------------------------|------------|------------|------------|---|------------------------------|------------------------------|----------|
|              |   |     |    |    |               |                      |                            |            |            |            | ER-, PR-, AR-, HER2-, BRAF V600E-, EBV- (ISH)                                   |                              |                              |          |
| <b>CB103</b> | M |     | 81 | 50 | retrospective | lymph node           | large cell carcinoma       | <b>POS</b> | <b>POS</b> | <b>POS</b> | CK14+, MUC1+, TTF1-, CD10-, CD117-, CD56-, Ki67 50%, EBV- (ISH)                 | lung                         | ND                           | LUAD     |
| <b>CB104</b> | M |     | 43 | 90 | retrospective | prostate             | squamous cell carcinoma    | <b>POS</b> | ND         | <b>NEG</b> | PSA-, P40+  | prostate or bladder          | bladder                      | LUSC     |
| <b>CB105</b> | M | E01 | 61 | 40 | retrospective | liver                | adenocarcinoma             | ND         | <b>POS</b> | <b>NEG</b> | CDX2-, MUC1+, MUC2+, TTF1-, HER2-   | gastrointestinal             | gastrointestinal             | STAD-CRC |
| <b>CB106</b> | M | E02 | 61 | 70 | retrospective | thyroid, post-mortem | neuroendocrine             | ND         | <b>POS</b> | <b>NEG</b> | CD56-, CDX2-, chromogranin +, MUC1+, MUC2+, synaptophysin -, TTF1-              | gastrointestinal             | gastrointestinal             | STAD-CRC |
| <b>CB108</b> | M | F01 | 74 | 80 | retrospective | bone                 | adenocarcinoma             | <b>POS</b> | <b>POS</b> | <b>NEG</b> | PSA-, TTF1-, CD34-  | upper gastrointestinal tract | upper gastrointestinal tract | STAD-CRC |
| <b>CB109</b> | M | F02 | 74 | 65 | retrospective | dermis               | undifferentiated carcinoma | <b>POS</b> | <b>POS</b> | <b>NEG</b> | CD31-, CD34-, CK 5/6-, Factor VIII-, LCA-, PSA-, S100-, SOX9-, TTF1-, vimentin- | upper gastrointestinal tract | ND                           | STAD-CRC |

|              |   |     |    |    |               |                         |                |            |            |            |   |                          |      |                 |
|--------------|---|-----|----|----|---------------|-------------------------|----------------|------------|------------|------------|---|--------------------------|------|-----------------|
| <b>CB110</b> | F |     | 57 | 70 | retrospective | brain                   | adenocarcinoma | <b>POS</b> | <b>POS</b> | <b>NEG</b> | BRAF V600E-, TTF1-                            | mullerian                | lung | OV              |
| <b>CB112</b> | M |     | 79 | 35 | retrospective | colon                   | adenocarcinoma | <b>POS</b> | <b>POS</b> | <b>NEG</b> | CDX2-, calretinin-, CDX2-, CEA-, PDPN-, TTF1- | ND                       | ND   | LUAD            |
| <b>CB115</b> | F |     | 86 | 50 | retrospective | muscle, post-mortem     | adenocarcinoma | <b>POS</b> | <b>POS</b> | <b>NEG</b> | TTF1-, CDX2-                                  | Intrahepatic biliar duct | ND   | STAD-CRC / PAAD |
| <b>CB116</b> | F |     | 65 | 60 | retrospective | ND, post-mortem         | adenocarcinoma | ND         | <b>POS</b> | <b>NEG</b> | TTF1-, CDX2-                                  | extrahepatic biliar duct | ND   | STAD-CRC / PAAD |
| <b>CB117</b> | M |     | 69 | 80 | retrospective | lymph node              | carcinoma      | <b>POS</b> | ND         | ND         | HMB45-, MART1-, S100-, CD10-                  | ND                       | ND   | LUAD            |
| <b>CB118</b> | M |     | 61 | 60 | retrospective | lymph node, post-mortem | adenocarcinoma | ND         | <b>POS</b> | <b>NEG</b> | CDX2-, TTF1-                                  | pancreas                 | lung | LUAD            |
| <b>CB119</b> | M | Q01 | 66 | 50 | retrospective | lymph node              | adenocarcinoma | <b>POS</b> | ND         | ND         | CK 5/6-, Ki67 50%, MUC1+, PSA-                | prostate or bladder      | ND   | HNSC / LUAD     |
| <b>CB120</b> | M | Q02 | 66 | 50 | retrospective | lymph node              | adenocarcinoma | <b>POS</b> | ND         | ND         | CK 5/6-, Ki67 50%, MUC1+, PSA-                | prostate or bladder      | ND   | HNSC / LUAD     |

|              |   |     |    |    |               |            |  |     |     |     |                                 |                  |                          |             |
|--------------|---|-----|----|----|---------------|------------|--|-----|-----|-----|---------------------------------|------------------|--------------------------|-------------|
| <b>CB121</b> | M | R01 | 69 | 30 | retrospective | bone       | adenocarcinoma   | POS | NEG | POS | BRAF V600E-, CDX2+, PSA-, TTF1- | gastrointestinal | bile duct                | CHOL / PAAD |
| <b>CB122</b> | M | R02 | 69 | 65 | retrospective | liver      | adenocarcinoma   | POS | NEG | POS | BRAF V600E-, CDX2+, PSA-, TTF1- | gastrointestinal | bile duct                | CHOL / PAAD |
| <b>CB125</b> | M |     | 64 | 60 | prospective   | cerebellum | mucinous adenocarcinoma  | ND  | NEG | ND  | PDL1-, TTF1-                    | ND               | gastrointestinal or lung | STAD-CRC    |
| <b>PF005</b> | F |     | 75 | ND | retrospective | lung       | poorly differentiated adenocarcinoma                                 | ND  | ND  | ND  | ND                              | ND               | ND                       | BRCA        |
| <b>PF006</b> | F |     | 81 | ND | retrospective | brain      | clear-cell carcinoma   | ND  | ND  | ND  | ND                              | kidney           | lung                     | LUAD        |
| <b>PF007</b> | M |     | 53 | ND | retrospective | liver      | adenocarcinoma   | ND  | ND  | ND  | ND                              | gastrointestinal | ND                       | STAD-CRC    |
| <b>PF011</b> | M |     | 75 | ND | retrospective | lung       | carcinoma with a transitional/squamous and glandular differentiation | ND  | ND  | ND  | TTF1-                           | ND               | ND                       | LUSC        |
| <b>PF013</b> | F |     | 71 | ND | retrospective | liver      | poorly differentiated adenocarcinoma                                 | ND  | ND  | ND  | ND                              | ND               | esophagus                | CHOL        |

|              |   |  |    |    |               |        |                |    |            |    |                       |          |                  |             |
|--------------|---|--|----|----|---------------|--------|----------------|----|------------|----|-----------------------|----------|------------------|-------------|
| <b>PF017</b> | M |  | 72 | ND | retrospective | liver  | adenocarcinoma | ND | ND         | ND | ND                    | ND       | gallbladder      | CHOL / PAAD |
| <b>PF018</b> | F |  | 79 | ND | retrospective | liver  | adenocarcinoma | ND | ND         | ND | ND                    | pancreas | small intestine  | STAD-CRC    |
| <b>PF019</b> | M |  | 74 | ND | retrospective | liver  | adenocarcinoma | ND | ND         | ND | ND                    | pancreas | pancreas         | PAAD        |
| <b>PF020</b> | M |  | 72 | ND | retrospective | pleura | adenocarcinoma | ND | ND         | ND | ND                    | ND       | lung             | PAAD        |
| <b>PF021</b> | F |  | 79 | ND | retrospective | pleura | adenocarcinoma | ND | <b>POS</b> | ND | TTF1+, CEA+, ER-, PR- | lung     | NA               | PAAD /BRCA  |
| <b>PF022</b> | M |  | 73 | ND | retrospective | pleura | adenocarcinoma | ND | ND         | ND | CEA+, TTF1-           | ND       | gastrointestinal | PAAD        |
| <b>PF024</b> | M |  | 77 | ND | retrospective | pleura | adenocarcinoma | ND | ND         | ND | TTF1+                 | lung     | NA               | LUSC        |
| <b>PF025</b> | M |  | 50 | ND | retrospective | pleura | adenocarcinoma | ND | ND         | ND | ND                    | ND       | lung             | LIHC        |

|              |   |  |    |    |               |            |                                 |            |                |            |  |                          |    |                 |
|--------------|---|--|----|----|---------------|------------|---------------------------------|------------|----------------|------------|--|--------------------------|----|-----------------|
| <b>PF059</b> | F |  | 73 | ND | retrospective | lymph node | poorly differentiated carcinoma | <b>POS</b> | <b>POS/NEG</b> | <b>NEG</b> | TTF1 -, P63-, CD45 -, CDX2 -, MART1-, S100 -, ER+/-, PR -, HER2-, Ki67 85%, EBV- (ISH) | breast                   | ND | STAD-CRC        |
| <b>PF080</b> | F |  | 52 | ND | retrospective | lymph node | adenocarcinoma                  | ND         | <b>POS</b>     | <b>POS</b> | CDX2+, ER+, PR+, CA125-, CD10-, CEA+, chromogranin -, NSE-, TTF1-, vimentin-, WT1-     | breast or genital system | ND | STAD-CRC / PAAD |

**Table 3. Association of miRNA expression with overall survival (significant miRNAs).**

| <b>miRNA</b>  | <b>HR</b> | <b>lower 95%</b> | <b>upper 95%</b> | <b>p-value</b> |
|---------------|-----------|------------------|------------------|----------------|
| miR-124-3p    | 0.11      | 0.03             | 0.36             | <b>0.00</b>    |
| miR-21-5p     | 7.00      | 2.10             | 23.00            | <b>0.00</b>    |
| miR-9-3p      | 0.29      | 0.12             | 0.71             | <b>0.01</b>    |
| miR-149-5p    | 0.32      | 0.13             | 0.78             | <b>0.01</b>    |
| miR-372-3p    | 0.33      | 0.12             | 0.89             | <b>0.03</b>    |
| miR-485-5p    | 0.37      | 0.16             | 0.90             | <b>0.03</b>    |
| miR-375       | 9.60      | 1.30             | 73.00            | <b>0.03</b>    |
| miR-25-3p     | 0.26      | 0.08             | 0.87             | <b>0.03</b>    |
| miR-27b-3p    | 2.60      | 1.10             | 6.10             | <b>0.03</b>    |
| miR-181a-2-3p | 0.38      | 0.15             | 0.93             | <b>0.03</b>    |
| miR-10b-5p    | 0.35      | 0.13             | 0.93             | <b>0.04</b>    |
| miR-96-5p     | 2.50      | 1.00             | 6.20             | <b>0.04</b>    |
| miR-423-5p    | 3.50      | 1.00             | 12.00            | <b>0.04</b>    |
| miR-214-3p    | 2.60      | 1.00             | 6.60             | <b>0.05</b>    |

For each miRNA the hazard ratio (HR) with 95% confidence interval and *p*-value (*p*) is reported for OS.

## References

1. Fizazi K, Greco FA, Pavlidis N, Daugaard G, Oien K, Pentheroudakis G & Committee EG (2015) Cancers of unknown primary site: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 26 Suppl 5, v133-138, doi: 10.1093/annonc/mdv305.
2. Pentheroudakis G, Briasoulis E & Pavlidis N (2007) Cancer of unknown primary site: missing primary or missing biology? *Oncologist* 12, 418-425, doi: 10.1634/theoncologist.12-4-418.
3. Rassy E & Pavlidis N (2019) The currently declining incidence of cancer of unknown primary. *Cancer Epidemiol* 61, 139-141, doi: 10.1016/j.canep.2019.06.006.
4. Briasoulis E, Fountzilas G, Bamias A, Dimopoulos MA, Xiros N, Aravantinos G, Samantas E, Kalofonos H, Makatsoris T, Mylonakis N, Papakostas P, Skarlos D, Varthalitis I & Pavlidis N (2008) Multicenter phase-II trial of irinotecan plus oxaliplatin [IROX regimen] in patients with poor-prognosis cancer of unknown primary: a hellenic cooperative oncology group study. *Cancer Chemother Pharmacol* 62, 277-284, doi: 10.1007/s00280-007-0604-7.
5. Pentheroudakis G, Briasoulis E, Kalofonos HP, Fountzilas G, Economopoulos T, Samelis G, Koutras A, Karina M, Xiros N, Samantas E, Bamias A, Pavlidis N & Hellenic Cooperative Oncology G (2008) Docetaxel and carboplatin combination chemotherapy as outpatient palliative therapy in carcinoma of unknown primary: a multicentre Hellenic Cooperative Oncology Group phase II study. *Acta Oncol* 47, 1148-1155, doi: 10.1080/02841860701843043.
6. Warner E, Goel R, Chang J, Chow W, Verma S, Dancy J, Franssen E, Dulude H, Girouard M, Correia J & Gallant G (1998) A multicentre phase II study of carboplatin and prolonged oral etoposide in the treatment of cancer of unknown primary site (CUPS). *Br J Cancer* 77, 2376-2380, doi: 10.1038/bjc.1998.395.
7. Gross-Goupil M, Fourcade A, Blot E, Penel N, Negrier S, Culine S, Chaigneau L, Lesimple T, Priou F, Lortholary A, Kaminsky MC, Provencal J, Voog E, Bouzy J, Laplanche A & Fizazi K (2012) Cisplatin alone or combined with gemcitabine in carcinomas of unknown primary: results of the randomised GEFCAPI 02 trial. *Eur J Cancer* 48, 721-727, doi: 10.1016/j.ejca.2012.01.011.
8. Hainsworth JD, Daugaard G, Lesimple T, Hubner G, Greco FA, Stahl MJ, Buschenfelde CM, Allouache D, Penel N, Knoblauch P & Fizazi KS (2015) Paclitaxel/carboplatin with or without belinostat as empiric first-line treatment for patients with carcinoma of unknown primary site: A randomized, phase 2 trial. *Cancer* 121, 1654-1661, doi: 10.1002/cncr.29229.
9. Huebner G, Link H, Kohne CH, Stahl M, Kretzschmar A, Steinbach S, Folprecht G, Bernhard H, Al-Batran SE, Schoffski P, Burkart C, Kullmann F, Otremba B, Menges M, Hoffmann M, Kaiser U, Aldaoud A & Jahn A (2009) Paclitaxel and carboplatin vs gemcitabine and vinorelbine in patients with adeno- or undifferentiated carcinoma of unknown primary: a randomised prospective phase II trial. *Br J Cancer* 100, 44-49, doi: 10.1038/sj.bjc.6604818.
10. Clynick B, Dessauvage B, Sterrett G, Harvey NT, Allcock RJN, Saunders C, Erber W & Meehan K (2018) Genetic characterisation of molecular targets in carcinoma of unknown primary. *J Transl Med* 16, 185, doi: 10.1186/s12967-018-1564-x.
11. Ross JS, Wang K, Gay L, Otto GA, White E, Iwanik K, Palmer G, Yelensky R, Lipson DM, Chmielecki J, Erlich RL, Rankin AN, Ali SM, Elvin JA, Morosini D, Miller VA & Stephens PJ (2015) Comprehensive Genomic Profiling of Carcinoma of Unknown Primary Site: New Routes to Targeted Therapies. *JAMA Oncol* 1, 40-49, doi: 10.1001/jamaoncol.2014.216.
12. Varghese AM, Arora A, Capanu M, Camacho N, Won HH, Zehir A, Gao J, Chakravarty D, Schultz N, Klimstra DS, Ladanyi M, Hyman DM, Solit DB, Berger MF & Saltz LB (2017) Clinical and molecular characterization of patients with cancer of unknown primary in the modern era. *Ann Oncol* 28, 3015-3021, doi: 10.1093/annonc/mdx545.
13. Horlings HM, van Laar RK, Kerst JM, Helgason HH, Wesseling J, van der Hoeven JJ, Warmoes MO, Floore A, Witteveen A, Lahti-Domenici J, Glas AM, Van't Veer LJ & de Jong D (2008) Gene expression profiling to identify the histogenetic origin of metastatic adenocarcinomas of unknown primary. *J Clin Oncol* 26, 4435-4441, doi: 10.1200/JCO.2007.15.4435 [pii]

10.1200/JCO.2007.14.6969.

14. Tothill RW, Shi F, Paiman L, Bedo J, Kowalczyk A, Mileskin L, Buela E, Klupacs R, Bowtell D & Byron K (2015) Development and validation of a gene expression tumour classifier for cancer of unknown primary. *Pathology* 47, 7-12, doi: 10.1097/PAT.000000000000194.

15. Varadhachary GR, Spector Y, Abbruzzese JL, Rosenwald S, Wang H, Aharonov R, Carlson HR, Cohen D, Karanth S, Macinkas J, Lenzi R, Chajut A, Edmonston TB & Raber MN (2011) Prospective gene signature study using microRNA to identify the tissue of origin in patients with carcinoma of unknown primary. *Clin Cancer Res* 17, 4063-4070, doi: 10.1158/1078-0432.CCR-10-2599.

16. Varadhachary GR, Talantov D, Raber MN, Meng C, Hess KR, Jatko T, Lenzi R, Spigel DR, Wang Y, Greco FA, Abbruzzese JL & Hainsworth JD (2008) Molecular profiling of carcinoma of unknown primary and correlation with clinical evaluation. *J Clin Oncol* 26, 4442-4448, doi: 10.1200/JCO.2007.14.4378 [pii]

10.1200/JCO.2007.14.4378.

17. Ferracin M, Pedriali M, Veronese A, Zagatti B, Gafà R, Magri E, Lunardi M, Munerato G, Querzoli G, Maestri I, Ulazzi L, Nenci I, Croce CM, Lanza G, Querzoli P & Negrini M (2011) MicroRNA profiling for the identification of cancers with unknown primary tissue-of-origin. *Journal of Pathology* 225, 43-53, Article, doi: 10.1002/path.2915.

18. Pentheroudakis G, Pavlidis N, Fountzilias G, Krikelis D, Goussia A, Stoyianni A, Sanden M, St Cyr B, Yerushalmi N, Benjamin H, Meiri E, Chajut A, Rosenwald S, Aharonov R & Spector Y (2013) Novel microRNA-based assay demonstrates 92% agreement with diagnosis based on clinicopathologic and management data in a cohort of patients with carcinoma of unknown primary. *Mol Cancer* 12, 57, doi: 10.1186/1476-4598-12-57.

19. Moran S, Martinez-Cardus A, Sayols S, Musulen E, Balana C, Estival-Gonzalez A, Moutinho C, Heyn H, Diaz-Lagares A, de Moura MC, Stella GM, Comoglio PM, Ruiz-Miro M, Matias-Guiu X, Pazo-Cid R, Anton A, Lopez-Lopez R, Soler G, Longo F, Guerra I, Fernandez S, Assenov Y, Plass C, Morales R, Carles J, Bowtell D, Mileskin L, Sia D, Tothill R, Tabernero J, Llovet JM & Esteller M (2016) Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol* 17, 1386-1395, doi: 10.1016/S1470-2045(16)30297-2.

20. Hainsworth JD & Anthony Greco F (2016) Lung Adenocarcinoma with Anaplastic Lymphoma Kinase (ALK) Rearrangement Presenting as Carcinoma of Unknown Primary Site: Recognition and Treatment Implications. *Drugs Real World Outcomes* 3, 115-120, doi: 10.1007/s40801-016-0064-7.

21. Hainsworth JD, Rubin MS, Spigel DR, Boccia RV, Raby S, Quinn R & Greco FA (2013) 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* 31, 217-223, doi: 10.1200/JCO.2012.43.3755.

22. Pentheroudakis G, Spector Y, Krikelis D, Kotoula V, Meiri E, Malamou-Mitsi V, Fountzilias G, Sanden M, Pavlidis N, Benjamin H & Aharonov R (2013) Global microRNA profiling in favorable prognosis subgroups of cancer of unknown primary (CUP) demonstrates no significant expression differences with metastases of matched known primary tumors. *Clin Exp Metastasis* 30, 431-439, doi: 10.1007/s10585-012-9548-3.

23. Yoon HH, Foster NR, Meyers JP, Steen PD, Visscher DW, Pillai R, Prow DM, Reynolds CM, Marchello BT, Mowat RB, Mattar BI, Erlichman C & Goetz MP (2016) Gene expression profiling identifies responsive patients with cancer of unknown primary treated with carboplatin, paclitaxel, and everolimus: NCCTG N0871 (alliance). *Ann Oncol* 27, 339-344, doi: 10.1093/annonc/mdv543.

24. Handorf CR, Kulkarni A, Grenert JP, Weiss LM, Rogers WM, Kim OS, Monzon FA, Halks-Miller M, Anderson GG, Walker MG, Pillai R & Henner WD (2013) A multicenter study directly comparing the diagnostic accuracy of gene expression profiling and immunohistochemistry for primary site identification in metastatic tumors. *Am J Surg Pathol* 37, 1067-1075, doi: 10.1097/PAS.0b013e31828309c4.

25. Weiss LM, Chu P, Schroeder BE, Singh V, Zhang Y, Erlander MG & Schnabel CA (2013) Blinded comparator study of immunohistochemical analysis versus a 92-gene cancer classifier in the diagnosis of the primary site in metastatic tumors. *J Mol Diagn* 15, 263-269, doi: 10.1016/j.jmoldx.2012.10.001

S1525-1578(12)00309-1 [pii].

26. Ettinger DS, Handorf CR, Agulnik M, Bowles DW, Cates JM, Cristea M, Dotan E, Eaton KD, Fidias PM, Gierada D, Gilcrease GW, Godby K, Iyer R, Lenzi R, Phay J, Rashid A, Saltz L, Schwab RB, Shulman LN,



- Smerage JB, Stevenson MM, Varadhachary GR, Zager JS, Zhen WK, Bergman MA & Freedman-Cass DA (2014) Occult primary, version 3.2014. *J Natl Compr Canc Netw* 12, 969-974, doi: 12/7/969 [pii].
27. Liu A, Tetzlaff MT, Vanbelle P, Elder D, Feldman M, Tobias JW, Sepulveda AR & Xu X (2009) MicroRNA expression profiling outperforms mRNA expression profiling in formalin-fixed paraffin-embedded tissues. *Int J Clin Exp Pathol* 2, 519-527.
28. Peiro-Chova L, Pena-Chilet M, Lopez-Guerrero JA, Garcia-Gimenez JL, Alonso-Yuste E, Burgues O, Lluch A, Ferrer-Lozano J & Ribas G (2013) High stability of microRNAs in tissue samples of compromised quality. *Virchows Arch* 463, 765-774, doi: 10.1007/s00428-013-1485-2.
29. Doleshal M, Magotra AA, Choudhury B, Cannon BD, Labourier E & Szafranska AE (2008) Evaluation and validation of total RNA extraction methods for microRNA expression analyses in formalin-fixed, paraffin-embedded tissues. *J Mol Diagn* 10, 203-211, doi: 10.2353/jmoldx.2008.070153  
S1525-1578(10)60151-1 [pii].
30. Li J, Smyth P, Flavin R, Cahill S, Denning K, Aherne S, Guenther SM, O'Leary JJ & Sheils O (2007) Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. *BMC Biotechnol* 7, 36, doi: 10.1186/1472-6750-7-36.
31. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, Knudsen BS, Stirewalt DL, Gentleman R, Vessella RL, Nelson PS, Martin DB & Tewari M (2008) Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* 105, 10513-10518, doi: 10.1073/pnas.0804549105  
0804549105 [pii].
32. Meiri E, Mueller WC, Rosenwald S, Zepeniuk M, Klinke E, Edmonston TB, Werner M, Lass U, Barshack I, Feinmesser M, Huszar M, Fogt F, Ashkenazi K, Sanden M, Goren E, Dromi N, Zion O, Burnstein I, Chajut A, Spector Y & Aharonov R (2012) A second-generation microRNA-based assay for diagnosing tumor tissue origin. *Oncologist* 17, 801-812, doi: 10.1634/theoncologist.2011-0466  
theoncologist.2011-0466 [pii].
33. Garajová I, Funel N, Fiorentino M, Agostini V, Ferracin M, Negrini M, Frassinetti GL, Gavelli G, Frampton AE, Biasco G & Giovannetti E (2015) MicroRNA profiling of primary pulmonary enteric adenocarcinoma in members from the same family reveals some similarities to pancreatic adenocarcinoma—A step towards personalized therapy. *Clinical Epigenetics* 7, Article, doi: 10.1186/s13148-015-0162-5.
34. Laprovitera N, Grzes M, Porcellini E & Ferracin M (2018) Cancer Site-Specific Multiple microRNA Quantification by Droplet Digital PCR. *Front Oncol* 8, 447, doi: 10.3389/fonc.2018.00447.
35. Ferracin M, Lupini L, Salamon I, Saccenti E, Zanzi MV, Rocchi A, Da Ros L, Zagatti B, Musa G, Bassi C, Mangolini A, Cavallesco G, Frassoldati A, Volpato S, Carcoforo P, Hollingsworth AB & Negrini M (2015) Absolute quantification of cell-free microRNAs in cancer patients. *Oncotarget* 6, 14545-14555, doi: 10.18632/oncotarget.3859.
36. Ferracin M, Salamon I, Lupini L, Miotto E, Sabbioni S & Negrini M (2016) Circulating microRNA quantification using DNA-binding dye chemistry and droplet digital PCR. *Journal of Visualized Experiments* 2016, Article, doi: 10.3791/54102.
37. Tibshirani R, Hastie T, Narasimhan B & Chu G (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 99, 6567-6572, doi: 10.1073/pnas.082099299  
99/10/6567 [pii].
38. Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B (Methodological)* 58, 267-288.
39. Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M, Benjamin H, Shabes N, Tabak S, Levy A, Lebanony D, Goren Y, Silberschein E, Targan N, Ben-Ari A, Gilad S, Sion-Vardy N, Tobar A, Feinmesser M, Kharenko O, Nativ O, Nass D, Perelman M, Yosepovich A, Shalmon B, Polak-Charcon S, Fridman E, Avniel A, Bentwich I, Bentwich Z, Cohen D, Chajut A & Barshack I (2008) MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol* 26, 462-469, doi: nbt1392 [pii]

10.1038/nbt1392.

40. Pentheroudakis G, Golfinopoulos V & Pavlidis N (2007) Switching benchmarks in cancer of unknown primary: from autopsy to microarray. *Eur J Cancer* 43, 2026-2036, doi: S0959-8049(07)00516-3 [pii]

10.1016/j.ejca.2007.06.023.

41. Sokilde R, Vincent M, Moller AK, Hansen A, Hoiby PE, Blondal T, Nielsen BS, Daugaard G, Moller S & Litman T (2014) Efficient identification of miRNAs for classification of tumor origin. *J Mol Diagn* 16, 106-115, doi: 10.1016/j.jmoldx.2013.10.001.

42. De Carolis S, Storci G, Ceccarelli C, Savini C, Gallucci L, Sansone P, Santini D, Seracchioli R, Taffurelli M, Fabbri F, Romani F, Compagnone G, Giuliani C, Garagnani P, Bonafe M & Cricca M (2019) HPV DNA Associates With Breast Cancer Malignancy and It Is Transferred to Breast Cancer Stromal Cells by Extracellular Vesicles. *Front Oncol* 9, 860, doi: 10.3389/fonc.2019.00860.

43. Otter S, Whitaker S, Chatterjee J & Stewart A (2019) The Human Papillomavirus as a Common Pathogen in Oropharyngeal, Anal and Cervical Cancers. *Clin Oncol (R Coll Radiol)* 31, 81-90, doi: 10.1016/j.clon.2018.10.004.

44. Piana AF, Sotgiu G, Muroli MR, Cossu-Rocca P, Castiglia P & De Miglio MR (2014) HPV infection and triple-negative breast cancers: an Italian case-control study. *Virol J* 11, 190, doi: 10.1186/s12985-014-0190-3.

45. Conway AM, Mitchell C, Kilgour E, Brady G, Dive C & Cook N (2019) Molecular characterisation and liquid biomarkers in Carcinoma of Unknown Primary (CUP): taking the 'U' out of 'CUP'. *Br J Cancer* 120, 141-153, doi: 10.1038/s41416-018-0332-2.

46. Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, Vessella RL & Tewari M (2013) Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat Methods* 10, 1003-1005, doi: 10.1038/nmeth.2633

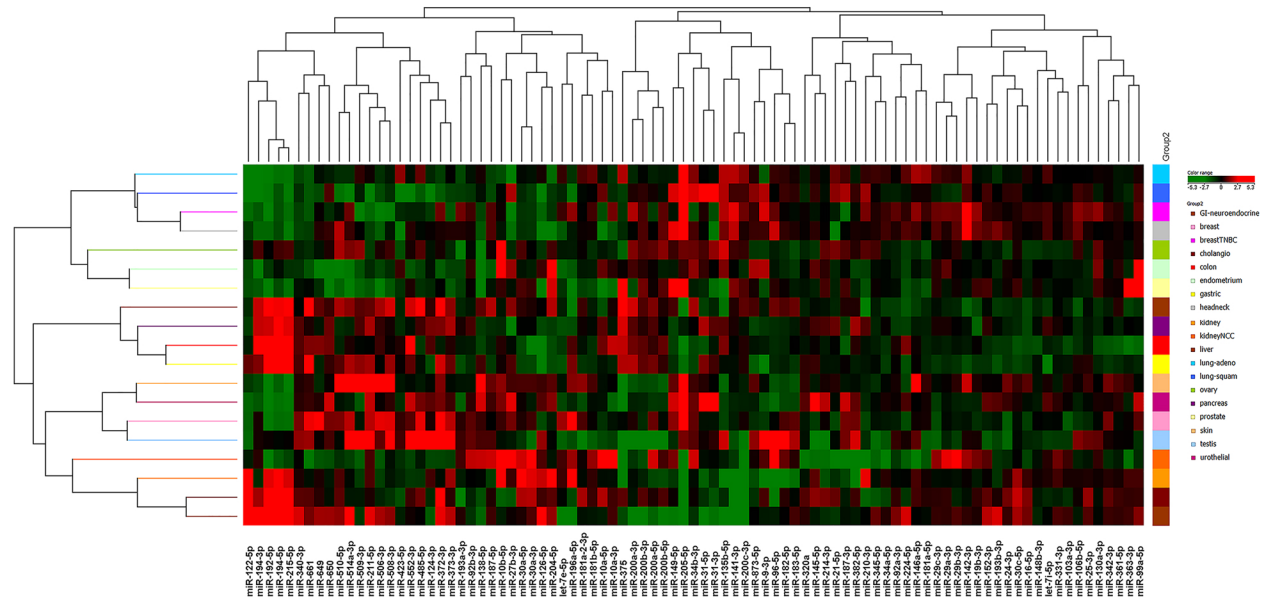
nmeth.2633 [pii].

47. Miotto E, Saccenti E, Lupini L, Callegari E, Negrini M & Ferracin M (2014) Quantification of circulating miRNAs by droplet digital PCR: comparison of EvaGreen- and TaqMan-based chemistries. *Cancer Epidemiol Biomarkers Prev* 23, 2638-2642, doi: 10.1158/1055-9965.EPI-14-0503

23/12/2638 [pii].

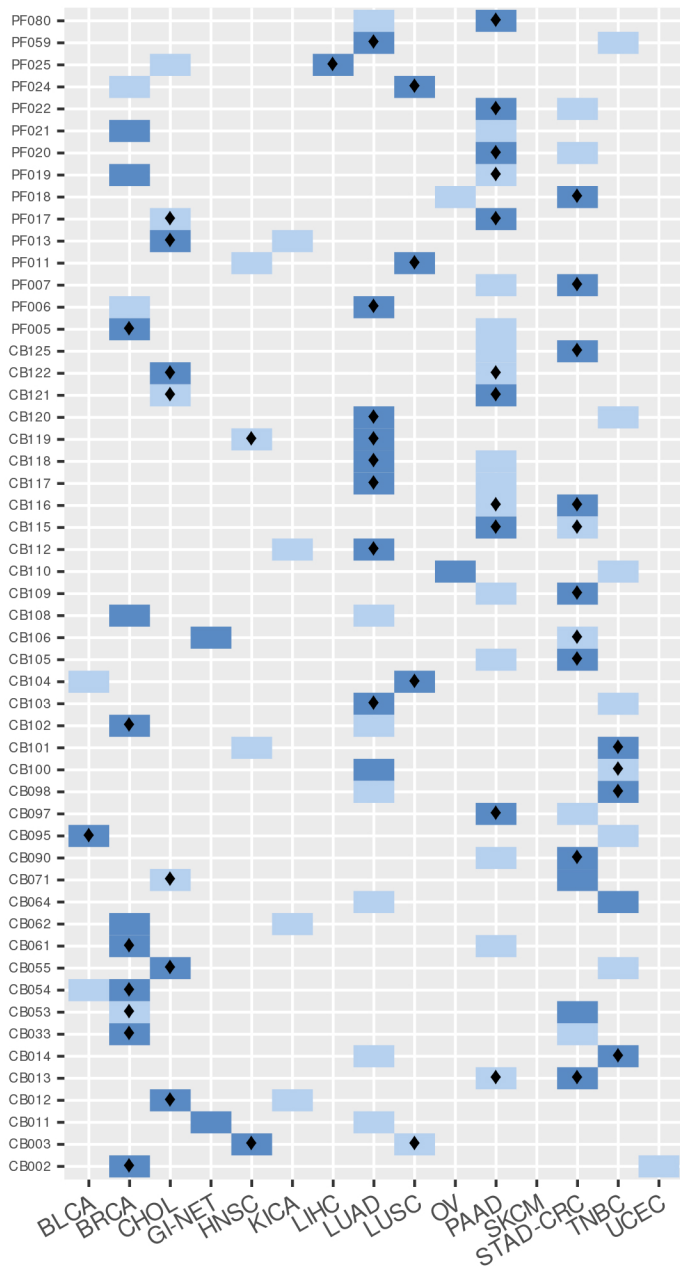
48. Zhang Y, Wang R, Li Y, Pan Y, Hu H, Zhang Y, Li H, Shen L, Yu Y, Sun Y & Chen H (2015) Negative Thyroid Transcription Factor 1 Expression Defines an Unfavorable Subgroup of Lung Adenocarcinomas. *J Thorac Oncol* 10, 1444-1450, doi: 10.1097/jto.0000000000000626.

# Figure 1

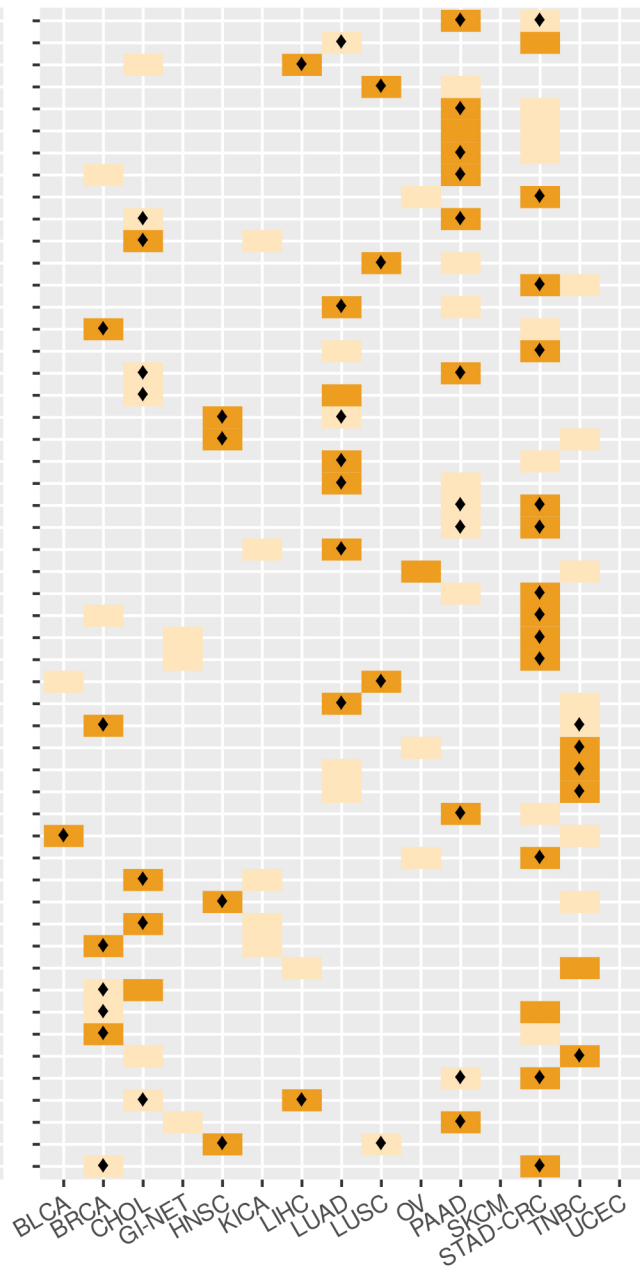


# Figure 2

## PAMR prediction



## Lasso prediction



# Figure 3

