ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

ONCOLOGIA, EMATOLOGIA E PATOLOGIA

Ciclo XXXVI

Settore Concorsuale: 06/A4 – Anatomia Patologica

Settore Scientifico Disciplinare: MED/08 – Anatomia Patologica

CANCERS OF UNKNOWN PRIMARY SITE (CUPs) AS A MODEL OF METASTATIC PROCESS.

Presentata da Mattia Riefolo

Coordinatore Dottorato

Chiar.ma Prof.ssa Manuela Ferracin

Supervisore

Chiar.ma Prof.ssa Antonietta D'Errico

Co-supervisore

Chiar.ma Prof.ssa Maria Pia Foschini

Esame finale anno 2023

INDEX

ABSTRACT

Cancers of unknown primary site (CUPs) are a rare group of metastatic tumours, with a frequency of 3-5%, but, with a median overall survival of 6-10 month, ranks among the ten most common causes of cancer-related deaths. The identification of a tumour primary site is usually reached by a combination of diagnostic investigations including physical examinations, blood analyses, imaging and immunohistochemical testing of the tumour tissue. In CUP patients, these investigations are inconclusive.

Since international guidelines for tumour treatment are based on primary site indication, CUP treatment requires a blind approach, which is very challenging for the oncologists. As a consequence, CUPs are usually empiric treated with poorly effective.

In this study, we applied a pre-determined set of microRNAs (miRNAs) using EvaGreen-based Droplet Digital PCR in a retrospective and prospective collection of formalin-fixed paraffin-embedded tissue samples. We assessed miRNA expression of 155 samples including primary tumours (N=94), metastases of known origin (N=10) and metastases of unknown origin (N=50). Then, we applied the shrunken centroids predictive algorithm to obtain the CUP's most probable site(s)-of-origin. The molecular test was successfully applied to all CUP samples and provided a site-of-origin identification (with a probability > 50%) for all samples, potentially within a one-week time frame from sample inclusion.

In the second part of the study we derived two CUP cell lines (CB055 and CB096), and corresponding patient-derived xenografts (PDXs), from ascites tumour cells. CUP cell lines and PDXs underwent histological, immune-phenotypical, molecular, and genomic characterization confirming the features of the original tumour. Tissues-of-origin prediction was obtained from the tumour microRNA expression profile and confirmed for CUP#55 by single cell RNA sequencing.

Genomic testing and FISH analysis identified FGFR2 amplification in both models. Drug-screening assays were performed to test the activity of FGFR2-targeting drug BGJ-398 (infigratinib) and the combination treatment with the MEK inhibitor trametinib, which proved to be synergic and exceptionally active, both in vitro and in vivo.

In conclusion, our study demonstrated that miRNA expression profiling in CUP samples could be employed as diagnostic test, concomitantly with the standard diagnostic workup, to offer valuable indication about the possible primary site thereby supporting treatment

decisions. Then we successfully derived two CUP models from patients, used for therapy tests, bringing personalized therapy closer to CUP patients and provides the rationale for FGFR2 targeting in metastatic tumours with FGFR2 pathway activation.

1 INTRODUCTION

1.1 Cancer of Unknown Primary Site

Cancer of Unknown Primary site (CUPs) are a heterogeneous group of metastatic tumours in which standardized diagnostic evaluation fails to identify the site of origin at the time of diagnosis (Fizazi et al. 2015).

The emergence of metastasis represents the main cause of cancer-related deaths (Chaffer and Weinberg 2011) and CUPs are the worst-case scenario of a metastatic disease.

CUPs patient has a median age of 60 years at presentation, with male prediction, poor prognosis and an unpredictable metastatic pattern and two or more organs with metastatic involvement at diagnosis (Golfinopoulos et al. 2009).

About 80% of CUP have a poor response to chemotherapy, with a median overall survival of 6-10 months, without any improving in the last decades (Fizazi et al. 2015; Massard, Loriot, and Fizazi 2011).

It is really difficult to provide incidence data about CUPs; historically representing 3–5% of worldwide cancer diagnoses (Fizazi et al. 2015), in the current era CUP reaches 1–2% (Massard, Loriot, and Fizazi 2011). This decreasing trend is probably due to the gradual improvement of diagnostic techniques that aid primary site detection, but socioeconomic disparities between countries seem to play an important role (Urban et al. 2013).

Despite the reduction in incidence CUP still ranks among the ten most common causes of cancer-related deaths (Kaaks et al. 2014; Pavlidis and Pentheroudakis 2012; Varadhachary and Raber 2014).

CUP biological entity to date remains an enigma, some researchers suggested that CUPs originate from small undetectable, dormant or later regressed primary lesions; others asserted that it is necessary to abandon the traditional tissue-gnostic approach and consider CUPs as early-disseminating, aggressive metastatic tumours with no existent primary site (Conway et al. 2019).

1.2 CUP Diagnosis

When dealing with a potential CUP diagnosis, clinical practice guidelines suggest a thorough diagnostic workup (Fizazi et al. 2015) which extensive physical examination, biochemical analyses and imaging tests (chest radiography and computed tomography (CT)

for thorax, abdomen and pelvis). If this initial series of diagnostic tests is still not able to identify the primary, additional specific tests can be considered according to metastases localization, patient's symptoms and gender; in these tests are included magnetic resonance imaging (MRI), endoscopy, positron emission tomography (PET) or serum tumor markers.

The last, and most critical step, is represented by immunohistochemical testing of tissue biopsy, which remains the most important diagnostic tool in establishing the tissue of origin. If the primary tumor remains elusive, a diagnosis of CUP is confirmed.

The most accurate - but not really useful to the single patient - diagnostic test is post-mortem autopsy. This is performed only in a small percentage of patients, but literature suggest that can unmask the primary site in70% of cases (Pentheroudakis, Golfinopoulos, and Pavlidis 2007). In particular, according to twelve cohort autopsy studies, five published between 1977-1980 (Didolkar et al. 1977; Nystrom et al. 1977; Snyder, Mavligit, and Valdivieso 1979; Stewart et al. 1979) and seven between 1981-2005 (Jordan and Shildt 1985; Hamilton and Langlands 1987; Le Chevalier et al. 1988; Maiche 1993; Mayordomo et al. 1993; Blaszyk, Hartmann, and Bjornsson 2003; Al-Brahim et al. 2005), most common primary sites identified post-mortem in CUP patients were lung (27%) and pancreas (24%), followed by hepatobiliary tree (liver and bile duct) (8%), kidneys or adrenals (8%), colon (7%) and genital organs (prostate 2%, testis 0,3%, ovary/uterus 4% and uterin cervix 0,7%) (Pentheroudakis, Golfinopoulos, and Pavlidis 2007).

Only one autopsy cohort was published 2022 (Vaideeswar, Patil, and Chaudhari 2022), but unfortunately it does not report the identification percentages of the primary site.

Moreover, many factors could lead to a wrong diagnosis of CUP, in particular:

1. A rushed premature diagnosis due to the clinician's insistence;
2. The inexperience or a wrong intuition of the pathologist;
3. Most important, the scarce availability of biological material that limits the number of performable diagnostic tests.

1.3 Histological and immunohistochemical evaluation

once the biopsy is obtained and the presence of malignancy is confirmed, a stepwise approach is used to identify the broad tumor type, the tumor subtype and, whenever possible, the site of origin. Since CUPs are prevalently carcinomas, first task of the pathologist is to evaluate, using a hematoxylin and eosin stained slide and routine light microscopy, the

basic form of differentiation. Most of them are adenocarcinoma (60%) or poorly differentiated carcinoma (30%), but they can also be squamous (5%) or neuroendocrine carcinoma (5%) (Oien 2009).

The underlying differences between distinct tissues depend on their gene expression profiles determined by regulation at multiple levels. Some active genes are ubiquitous since involved in basic cellular functions, other are expressed by one or few tissue types, impacting their function and differentiation (Ramskold et al. 2009). The expression of a subset of genes is retained in metastatic cells and it is used to trace back the tissue of origin.

However, this condition is better preserved in well-differentiated rather than poorly or undifferentiated cancers (Gevaert et al. 2009). Even though gene expression might be lower in metastases compared to primary tumors, it could provide a decisive clue to identify the tissue of origin; for this reason, immunohistochemical (IHC) testing covers an essential role in pathological evaluation of metastases.

The first IHC testing is made to exclude highly treatable tumor types such as lymphoma, melanoma, germ-cell tumors or sarcoma. When dealing with a carcinoma a first panel of 19 antibodies is assessed, followed by other additional testing if the previous were inconclusive (NCCN v.2/2023). Despite the existence of specific practice guidelines for IHC testing, a defined diagnostic algorithm to identify the tissue of origin, sometimes theory clashes with reality. In fact, when biological material available is scarce, pathologist's intuition and experience play a pivotal role in markers selection. For this reason, one of the major limits of this technique is the lack of objective evaluation parameters. Moreover, this process can be time-consuming, complex and can lead to unambiguous classification, especially when dealing with poorly differentiated tumors. Different cancer types can also share the expression of some markers, making really challenging to identify a univocal primary site.

Once establish the epithelial origin, the expression of two keratins (previously referred as cytokeratins), K7 and K20, allowing to broadly define the subset of carcinoma, are the most used for CUP primary site's prediction (Chu, Wu, and Weiss 2000). Belonging to the family protein of intermediate filaments, keratins are expressed in epithelial cells and derive from 54 functional genes in human genome (Schweizer et al. 2006). While K7 is found in some simple epithelia, K20 is expressed in a restricted setting of epithelia.

However, beyond this general rule, some carcinomas are notable for their lack of immunophenotypes: i.e. pancreatic carcinoma generally show K7+/K20+, even though can

lose K7; gastric adenocarcinoma can show all K7/K20 phenotypes; cholangiocarcinoma can overlap with pancreatic carcinoma (Duval, Savas, and Banner 2000; Shimonishi, Miyazaki, and Nakanuma 2000; Wildi et al. 1999), as shown in Table 1.

Once established the four K7/K20 expression pattern (K7+/K20+; K7+/K20-; K7-/K20+; K7-/K20-), IHC testing can proceed with site-specific markers.

K7+/K20+	K7+/K20-	K7-/K20+	K7-/K20-
Transitional cell ***Ptable ancreas***	Lung (adenocarcinoma) Breast Oviarian Endometrial Gastric Cholangiocarcinoma Thyroid	Colorectal	Hepatocellular carcinoma Lung (neuroendocrine) Renal Cell Prostate

Table 1. K7 and K20 expression in most common carcinoma.

1.4 CUP classifiation and treatment

International guidelines for tumor treatment are essentially based on primary site indication. With no evidence of a primary site, CUP patients cannot be treated with site-specific therapy, so they are managed based on their clinicopathological characteristics.

According to these criteria, they are classified into two prognostic subgroups, defined as favorable (15-20%) and unfavorable CUPs (80-85%). Favorable CUPs, present common features with a specific known tumor type, sharing its metastatic pattern, response to treatment and prognosis. Their life expectancy is longer (15-20 months) (Hainsworth and Fizazi 2009). An example of this subgroup are patients with cervical lymph node metastases from squamous cell carcinoma or with peritoneal carcinomatosis from papillary carcinoma (Hainsworth and Fizazi 2009).

Unfortunately, most CUP, belonging to the unfavorable group, with 6-10 months of median survival (Greco and Pavlidis 2009).

A metanalysis evaluated the impact on CUP survival of platinum, taxane or new-generation compounds (gemcitabine, vinca alkaloids or irinotecan) regimens, in monotherapy or in

combination; of note no chemotherapy agent was found to effectively prolong CUP patient survival (Golfinopoulos et al. 2009).

Nevertheless, in the NCCN guidelines are listed 11 chemotherapy regimens indicated for adenocarcinoma and 9 for squamous histology (NCCN v.2/2023). However, these guidelines remains empirical since are mostly based on single-arm phase II clinical trials (Briasoulis et al. 2008; Pentheroudakis et al. 2008; Warner et al. 1998) or small trials (Gross-Goupil et al. 2012; Hainsworth et al. 2015; Huebner et al. 2009).

1.5 Molecular prediction of the primary site

In the past decade, different approaches based on mRNA, microRNA or DNA methylation analysis were developed. Although based on different multi-feature algorithms, these molecular predictors use the same approach: the analysis of a cohort of tumors with known primary site (reference or training set) creates a cancer-specific signature database that is employed as a tissue-of-origin classifier; then, based on the similarities with the tumor types included in the reference set, a tissue-of-origin is assigned to each metastatic tumor with uncertain/unknown origin.

The accuracy of each molecular tool can be directly estimated by analyzing the same primary tumors used to create the classifier and independent datasets of primaries and metastatic tumors of known origin.

In general, molecular profiling tools developed to date declared an overall prediction accuracy that ranges between 80-95%, reflecting the great potential of these tools to assist and improve the diagnostic workup of CUP patients and offer them more therapeutic options (Losa et al. 2018).

The verification of prediction consistency in CUPs remains very challenging, since only in few cases post-mortem autopsy is performed and is not always sufficient to unmask the site-of-origin; at best the origin site is identified during the course of the disease, but in most cases prediction is evaluated using available surrogate measures (IHC, clinical presentation, response to therapy).

Due to the lack of robust evidences from randomized, prospective trials of a clinical benefit on CUPs from molecularly-directed therapies, to date international and national guidelines do not recommend the employment of tissue-of-origin predictors in their clinical management (Losa et al. 2018).

1.6 Gene expression profiling

Several researchers developed microarray-based gene-expression profiles (GEP) molecular tools able to classify human cancers (Golub et al. 1999; Ramaswamy et al. 2001; Su et al. 2001).

Of note, their first application aimed to the differential diagnosis of acute myeloid leukemia from acute lymphoblastic leukemia (Golub et al. 1999).

Later, several molecular tests were developed and commercialized, such as Pathwork Tissue of Origin Test® (Pathwork Diagnostics), to firstly explore their potential application for tissue of origin prediction in metastases of known origin (Dumur et al. 2008; Monzon et al. 2009; Pillai et al. 2011). In these microarray-based studies the prediction accuracy was around 85-95%, analyzing both frozen and FFPE samples, with poorly differentiated and undifferentiated primary tumors responsible for lower accuracy rates.

In addition, in a blinded-multicentric study, GEP prediction accuracy was compared with standard IHC in metastases of known origin. When the diagnosis was achieved with a single round of stains, IHC and GEP performance were similar, showing an accuracy higher than 90%; on the other hand, in those cases that required a second round of stains, GEP proved to be more accurate than IHC (91% v 71%; p=0.023) to determine the appropriate primary site (Handorf et al. 2013).

When applied on CUPs, in both retrospective and prospective settings, gene expression microarray (GEM) studies reported a consistency with clinical evaluation and pathological assessment that ranged from 62.5-85% (Horlings et al. 2008; Monzon et al. 2010; Tothill et al. 2015; Yoon et al. 2016).

However, microarray is a complex, time-consuming and expensive technology. For this reason, Ma et al. developed an RT-qPCR-based assay able to distinguish and classify 26 cancer types (Ma et al. 2006) with an overall sensitivity of 87% (Kerr et al. 2012).

This assay, commercially available since 2013 as CancerTypeID® (bioTheranostics), and is used in USA to solve uncertain cases that proved higher prediction accuracy than standard IHC (79% v 69%) (Hainsworth et al. 2013; Weiss et al. 2013).

In a recently published study, a machine learning method called SCOPE (Supervised Cancer Origin Prediction Using Expression) was proposed as a diagnostic tool able to incorporate information from whole-transcriptome RNA sequencing data and solve complex diagnoses. Training was performed on public datasets, The Cancer Genome Atlas (TCGA) included (10688 samples of 40 untreated primary tumor types and 26 adjacent normal tissues), while testing was performed retrospectively on untreated primary mesothelioma (N=211) and treatment-resistant metastatic cancers (N=201). SCOPE was able to achieve 99% of accuracy on mesotheliomas and 86% for sarcomatoid subgroup. This approach was also tested on 15 CUPs with results in accordance with conventional pathology in 80% of cases (Grewal et al. 2019).

In contrast, very few clinical trials and case reports evaluated the impact of molecular profiling on survival when treated with site-specific regimens according to molecular prediction (AC et al. 2015; Greco et al. 2015; Hainsworth et al. 2013; Varadhachary et al. 2008).

1.7 microRNA expression profiling

Formalin fixation followed by embedding in paraffin has been the method of choice to preserve tissue samples. However, both the fixation and embedding processes, significantly impact the RNA integrity and induce nucleotide chemical modifications (Masuda et al. 1999; von Ahlfen et al. 2007). Moreover, many other factors can impact RNA quality in FFPE samples: the time elapsed before the fixation, the thickness of the tissue, the temperature during fixation and the storage time of FFPE samples (von Ahlfen et al. 2007).

Due to their higher resistance to chemical modification or degradation over time, microRNAs can be robustly detected irrespective of the quality of the sample (Liu et al. 2009; Peiro-Chova et al. 2013).

Given the well-documented deregulation of miRNAs in cancer, the expression of specific miRNA signatures can efficiently discriminate tumor from normal tissues and different cancer types (Ferracin, Veronese, and Negrini 2010; Laprovitera et al. 2018).

Microarray-based studies explored the feasibility of miRNA-based detection of the tissues from which metastases arose showing high accuracy rates (78-90%), when dealing with metastases of known origin (Ferracin et al. 2011; Laprovitera, Riefolo, et al. 2021; Rosenfeld et al. 2008; Sokilde et al. 2014).

Rosenfeld et al. analyzed the profiles of 253 samples (up to 25 different histological subtypes) and built two classifiers by using decision-tree and K-nearest neighbors (KNN) algorithms (Rosenfeld et al. 2008). From this study a 48-miRNA signature proved high accuracy to reveal the correct tissue-of-origin on metastases of known origin. This molecular assay was translated from microarray to RT-qPCR technology, preserving its analytical validity (Rosenfeld et al. 2008; Rosenwald et al. 2010) to finally become the miRview mets®, which is an LDT in USA developed by Rosetta Genomics.

When applied on a prospective series of 74 CUPs, molecular prediction was found consistent with clinicopathological features in 84% of cases (Varadhachary et al. 2011). Moreover, in a retrospective study this assay was used to predict the tissue-of-origin of 57 patients, originally diagnosed as CUPs, characterized by central nervous system (CNS) metastases; predictions matched in 88% of cases with clinicopathological evaluation at the diagnosis, information obtained during the course of the disease or specific examinations performed after and based on the results of the molecular assay (Mueller et al. 2011).

To expand the number of tumor classes and include a wider range of carcinomas and neuroendocrine tumors, a second-generation assay named miRviews met2® (Rosetta Cancer Origin Test™) was designed, reaching up to 42 different tumor types and 64 miRNA. Custom microarray data on a training set of 1282 samples were used to build a classifier with a combination of two different algorithms, with ac accuracy of 85% (Meiri et al. 2012).

When applied on 84 CUP samples using this signature, molecular prediction was found concordant with the first clinical hypothesis in 70% of patients, with treatment response in 89% of cases and with the final clinical diagnosis according to supplemental IHC stains in 92% of patients (Pentheroudakis et al. 2013).

Another microarray-based molecular study identified a 47-miRNA signature able to discriminate ten different cancer types with a prediction accuracy of 100% in primary cancers and 78% in metastases of known origin, moreover, when applied to predict the primary site of 16 CUPs, predictions were found to be consistent with clinical and pathological hypotheses (Ferracin et al. 2011).

1.8 DNA methylation profiling

Given its role as regulatory mechanism of transcription, DNA methylation is one of the earliest explored epigenetic biomarkers of cancer. Acting as a stably inherited modification, it affects gene activity and cellular biology.

A comprehensive study with GoldenGate Assay by Illumina a DNA methylation fingerprint of 1628 human samples was obtained and 1505 CpG sites were interrogated. They analyzed 1054 tumorigenic samples, including 855 primary malignancies (19 solid tumor types and 10 types of hematologic malignancies), 50 metastatic tumors, 25 premalignant lesions, 82 cancer cell lines and 42 CUPs. Methylation patterns was found to be tissue-type specific and it was demonstrated that during tumorigenesis human cancer cells undergo a progressive gain of promoter CpG-island hypermethylation and a loss of CpG methylation in non-CpG-island promoters. Comparing DNA methylation fingerprints of CUPs with primary and metastatic tumors with known origin, they were able to assign a site-of-origin in 69% of CUP cases and create a prediction heatmap. Tumor type prediction was fully confirmed in 78% of cases in which pathological information was available (Fernandez et al. 2012).

Then, a classifier based on microarray DNA methylation signatures (EPICUP assay) using a training set of 2790 known primary tumors and 216 CUP samples, showed an overall accuracy of 90% when compared with the unmasked primary site (during the disease follow-up or in post-mortem autopsy (Moran et al. 2016). These results were in accordance to previous reports employing other molecular assays (Greco et al. 2015; Hainsworth et al. 2013; Yoon et al. 2016).

In a recent in-silico study, a combined miRNA- and methylation-based classifiers were trained in 6602 samples for miRNA-based profiles and a total of 5379 samples for DNA methylation-based profiles were collected from The Cancer Genome Atlas (TCGA), reaching an accuracy around 95% for DNA methylation and 91% for miRNA (Tang et al. 2018). However, this approach was not yet tested on CUPs.

Studies with tissue-of-origin profiling assays applied on CUP patients are summarized in Table 2.

Study Reference	Study Type	Data Type	No. of CUP	Type of Sample	Method Analysis	No. of Tumour Types	Tumour Type Included in the Reference Set	Most frequent Types	Validation	CUP Prediction Accuracy	Impact on Clinical Outcomes with Prediction-Based Treatment
(Tothill et al. 2005)	Retrospective	mRNA	13	FF and FFPE	10,500-gene GEM	14	breast, CRC, gastric, lung, melanoma, mesothelioma, ovarian, pancreas, prostate, renal, cSCC, head and neck, testicular, uterine	ND	Clinicopathological features	84%	ND
(Talantov et al. 2006)	Retrospective	mRNA	48	FFPE	10-gene RT-qPCR;	6	lung, breast, colon, ovary, pancreas, and prostate	ND	unmasked latent primary tumour sites identified months to years later from the initial diagnosis	76%	ND
(Horlings et al. 2008)	Retrospective	mRNA	38/38	FFPE	1900-gene GEM; CupPrint®(Agendia, Amsterdam, The Netherlands)	49	ND	lung (24%), CRC (18%), pancreas (16%), ovarian (11%)	Clinicopathological features and IHC	85%	ND
(Varadhachary et al. 2008)	Prospective/ retrospective	mRNA	104/120	FFPE	10-gene RT-qPCR	6	lung, breast, colon, ovary, pancreas, and prostate	CRC (49%), NSCLC (33%), pancreas (21%), ovarian (14%)	Clinicopathological features	61%	Improved OS of prospective cohort treated

											with site-specific regimen vs retrospective cohort treated with empiric chemotherapy
(Bridgewater et al. 2008)	Retrospective	mRNA	21	FFPE	1900-gene GEM; CupPrint®(Agendia, Amsterdam, The Netherlands)	49	ND	colon (19%), small and large bowel (19%), ovarian (19%), breast (9%)	Clinicopathological features	67%	ND
(Monzon et al. 2010)	Retrospective	mRNA	16/21	fresh-frozen	1550-gene GEM; Pathwork® TOO Test (Pathwork Diagnostic, Redwood City, CA, USA)	15	bladder, breast, colorectal, gastric, hepatocellular, kidney, non–small cell lung, ovarian, pancreatic, prostate, thyroid carcinomas, melanoma, testicular, germ cell tumour, non-Hodgkin's lymphoma, and sarcoma	CRC (5), breast (4), ovary (3), lung (2), and pancreas (2)	Clinicopathological features and IHC	62.5%	ND
(Ferracin et al. 2011)	Retrospective	miRNA	16	FFPE	microArray; 47-miRNA signature	10	breast, colon, endometrium, stomach/gastri	gastric (31%), lung (25%), pancreas	Clinicopathological features	85%	ND

							c, kidney, liver, lung, pancreas, prostate, and melanoma	(19%), liver (12%), colon (6%), and kidney (6%)			
(Mueller et al. 2011)	Retrospective	miRNA	57/60	FFPE	48-miRNA RT-qPCR; miRNA	25	ND	ND	Clinicopathological features	ND	ND
(Varadhachary et al. 2011)	Prospective	miRNA	74/104	FFPE	48-miRNA RT-qPCR; miRNA	25	ND	CRC (17%), ovarian (8%); SCC (8%), pancreaticobiliary (13%), NSCLC (8%)	Clinicopathological features	84%	ND
(Fernandez et al. 2012)	Retrospective	DNA methylation	42	fresh-frozen	DNA methylation array, 1505 CpG sites selected from 807 genes; GoldenGate assay (Illumina, San Diego, CA, USA)	29	bladder, breast, cervix, colon, endometrium, esophagus, ganglioneurom, glioma, head and neck, kidney, liver, melanoma, neuroblastoma, NSCLC, ovarian, pancreas, prostate, stomach, testis, ALL, AML, CLL, DLBCL, FL, MCL, mBL, MM, MDS/MPS, mixed lineage leukemia.	CRC (34%), NSCLC (17%), breast (17%)	Clinicopathological features	78%	ND

(Hainsworth et al. 2012)	Retrospective	mRNA	42	FFPE	92-gene RT-qPCR CancerTypeID®(bioTheranostics, San Diego, CA, USA)	30	CRC, Lung-adeno/large cell, breast, HCC, ovary, pancreas, kidney, bladder, gallbladder, skin/squamous, melanoma, sarcoma, endometrium, testis, thyroid, stomach, mesothelioma, prostate, brain, lymphoma, uterine	selection of CUP predicted as CRC	Clinicopathological features	54–86%	Improved OS compared to historical controls; better outcome in more responsive e predictive tumour types
(Greco et al. 2013)	Prospective/retrospective	mRNA	144/149	FFPE	92-gene RT-qPCR; CancerTypeID®(bioTheranostics, San Diego, CA, USA)	30	CRC, Lung-adeno/large cell, breast, HCC, ovary, pancreas, kidney, bladder, gallbladder, skin/squamous, melanoma, sarcoma, endometrium, testis, thyroid, stomach, mesothelioma, prostate, brain, lymphoma, uterine	CRC (15%), lung-adeno/large cell (10.5%), breast (8.8%), HCC (5.8%), ovary (5.2), pancreas (5.2%), kidney (4%), bladder (4%)	Clinicopathological features, IHC and unmasked latent primary tumour sites identified months to years later from the initial diagnosis	74–77%	ND

(Hainsworth et al. 2013)	Prospective	mRNA	252/289	FFPE	92-gene RT-qPCR CancerTypeID®(bioTheranostics, San Diego, CA, USA)	30	CRC, lung-adeno/large cell, breast, HCC, ovary, pancreas, kidney, bladder, gallbladder, skin/squamous, melanoma, sarcoma, endometrium, testis, thyroid, stomach, mesothelioma, prostate, brain, lymphoma, uterine	biliary tract (21%); CRC (10%); NSCLC (7%); breast (5%); urothelial (11%); pancreatic (5%)	ND	ND	Improved OS compared to historical
(Pentheroudakis et al. 2013)	Retrospective	miRNA	84/93	FFPE	64-miRNA RT-qPCR; miRviews met2 (Rosetta Genomics, Princeton, NJ, USA)	42	ND	ND	Clinicopathological features and IHC	92%	ND
(Greco et al. 2015)	Retrospective	mRNA	30/25	FFPE	92-gene RT-qPCR; CancerTypeID®(bioTheranostics, San Diego, CA, USA)	30	CRC, lung-adeno/large cell, breast, HCC, ovary, pancreas, kidney, bladder, gallbladder, skin/squamous, melanoma, sarcoma, endometrium, testis, thyroid, stomach, mesothelioma, prostate,	carcinoma (40%) (of which 30% were germ cell tumours and 20% were neuroendocrine tumours), sarcoma (30%), melanoma (20%) and lymphoma (8%)	Clinicopathological features and additional IHC or genetic testing post molecular profiling	84%	Improved PFS compared to historical

							brain, lymphoma, uterine				
(Tothill et al. 2015)	Retrospective	mRNA	49	FFPE	>25,000 genes, GEM	18	urinary (bladder), breast, cholangiocarcinoma, CRC, gastric, kidney, HCC, lung, neuroendocrine, ovary, pancreas, prostate, squamous cell, thyroid, melanoma, mesothelioma, sarcoma, testis	SCC (18%), lung (14%), kidney (10%), CRC (8%), breast (8%), cholangiocarcinoma (4%), mesothelioma (4%)	Clinicopathological features	78%	ND
(Moran et al. 2016)	Retrospective	DNA methylation	216	fresh-frozen	DNA methylation array, 485,577 CpG sites; EPICUP assay (FERRER, Barcelona, Spain)	38	ND	NSCLC (21%), head and neck SCC. (10%), breast (9%), colon (9%), HCC (7%), pancreas (7%)	Latent primary, clinicopathological features, autopsy	87–100%	Improved OS compared to patients treated with empiric chemotherapy
(Yoon et al. 2016)	Prospective phase II	mRNA	38/46	FFPE	2000-gene GEM; Pathwork® OO Test (Pathwork Diagnostic, Redwood City, CA, USA)	15	bladder, breast, colorectal, gastric, hepatocellular, kidney, non–small cell lung, ovarian, pancreatic, prostate, thyroid	NSCLC (21%), CRC (18%), ovary (18%), pancreas (16%)	ND	ND	Improved OS in platinum-responsive tumour types

							carcinomas, melanoma, testicular, germ cell tumor, non-Hodgkin's lymphoma, and sarcoma				
(Grewal et al. 2019)	Retrospective	mRNA	dic-15	FF and FFPE	Whole Transcriptome	40	ND	ND	Clinicopathological features	ND	ND
(Hayashi et al. 2019)	Prospective phase II	mRNA	130	FF	22,000-gene GEM; mRNA	39	ND				

Table 2. Tissue-of-origin profiling assays applied on CUP patients.

1.9 Driver mutations and precision medicine

Since our molecular understanding of cancer is rapidly expanding, histological cancer classification seems to be gradually replaced by a more informative molecular classification.

In fact, IMPACT study (Initiative for Molecular Profiling and Advanced Cancer Therapy), patients with advanced cancers received matched targeted therapies (MTT) based on their mutational profile. Results from this study strongly support the employment of genomic matching: MTT-treated patients showed higher response rates (complete and partial response) (11% v 5%; p = .0099), longer failure-free survival (3.4 v 2.9 months; p = .0015), and longer overall survival (8.4 v 7.3 months; p = .041). In addition, when treated with MTT, a significant difference was observed in OS between responders and nonresponders (23.4 v 8.5 months; p < .001) (Tsimberidou et al. 2017).

Therefore, since the identification of CUP-specific druggable alterations could address their limited treatment options, several researchers focused on the analysis of CUP mutational landscape.

High heterogeneity across different CUP patients was reported by several studies (Ross et al. 2015; Clynick et al. 2018; Varghese et al. 2017); of note, most common alterations harbored on driver genes mutated in many different cancer types, mostly affecting core mitogenic, cell growth pathways (especially PI3K and MAPK pathways) and epigenetic deregulation (Gatalica et al. 2014; Subbiah et al. 2017; Tothill et al. 2013). In a minority of CUP patients a common molecular signature was identified, mainly involving the exposure to exogenous mutagens, such as smoking or UV exposure (Tothill et al. 2013; Varghese et al. 2017).

Principal studies are summarized in Table 3.

Study	Analyzed Genes	Reference	No. of CUP Patients	% of Samples Alteration	% Potentially Targetable Alterations	Most Common Genetic Aberrations
Retrospective target sequencing and CNA analysis	701	(Tothill et al. 2013)	16	100	81	TP53 (62%), GNAS (25%), NOTCH1 (18%), NOTCH2 (18%), CDKN2A (18%), PIK3CA (18%), BRCA1 (18%), and STK11 (18%).
Retrospective metanalysis of mutational profiling, CNA analysis and protein expression profiling	47	(Gatalica et al. 2014)	1806	ND	96	TP53 (38%), KRAS (18%), BRCA2 (11%), PIK3CA (9%), STK11 (6%); EGFR (17%) and ERBB2 (5%) for amplification.
Retrospective targeted sequencing of CTNNB1, MET, PIK3CA, KRAS and BRAF	5	(Pentheroudakis et al. 2014)	87	66	37	CTNNB1 (19.5 %), KRAS (10.2%), PIK3CA (7%), BRAF (4.5%), and MET (4.5%).
Retrospective target sequencing using FoundationOne assay (Foundation Medicine, Cambridge, MA, USA)	236	(Ross et al. 2015)	200	96	20	TP53 (55%), KRAS (20%), CDKN2A (19%), MYC (12%), ARID1A (11%), MCL1 (10%), PIK3CA (9%), ERBB2 (8%), PTEN (7%), EGFR (6%), SMAD4 (7%), STK11 (7%), SMARCA4 (6%), RB1 (6%), RICTOR (6%), MLL2 (6%), BRAF (6%), and BRCA2 (6%).
Retrospective target sequencing using Oncomine Focus Assay and Cancer Hotspot v2 panel (Thermo Fisher Scientific, Waltham, MA, USA)	76	(Clynick et al. 2018)	21	81	52	TP53 (47%), KRAS (12%), MET (12%) and MYC (12%).
Retrospective target sequencing using MSK impact assay	468	(Varghese et al. 2017)	150	91	30	TP53 (70%), KRAS (35%), CDKN2A (30%), KEAP1 (23%), and SMARCA4 (22%)
Retrospective targeted sequencing and copy number analysis	50	(Loffler et al. 2016)	55	84	15	TP53 (55%), CDKN2A (22%), KRAS (18%), SMAD4 (11%), FGFR3 (9%), ATM (7%), and RB1 (7%).
Retrospective target sequencing using NGS CLIA-certified assay	-	(Subbiah et al. 2017)	17	88	41	Impaired P signaling (47%), Epigenetic deregulation (47%), and impaired cell cycle control (47%)
Retrospective target sequencing using NGS CLIA-certified assay	592	(Gatalica et al. 2018)	389	ND	22	TP53 (54%), KRAS (22%), ARID1A (13%), PIK3CA (9%), CDKN2A (8%), and SMARCA4 (7%)

Table 3. Studies analyzing the mutational landscape of CUP tumours.

Since immunotherapy is emerging as potentially winning therapeutic strategy in several cancer types, its possible administration on CUP patients have recently gained increasing interest. Immunotherapy biomarkers of response are currently intensively studied and to date include programmed cell death ligand 1 (PD-L1) expression, high tumor mutational burden (TMB) and high microsatellite instability (MSI) that have been reported to occur in 28% of CUP cases (Chalmers et al. 2017; Gatalica et al. 2014; Gatalica et al. 2018).

NCCN guidelines support the use of pembrolizumab, an anti-PD1 antibody, which is FDA-approved on MSI-H or dMMR patients regardless of the primary site (NCCN v.2/2023); in fact, this immune checkpoint inhibitor proved its antitumor activity in different cancer types.

To our knowledge two published case reports describe CUP patients (1 with high PDL1 expression and 1 with deficiency of the mismatch repair system or dMMR) that received immunotherapy treatment with promising outcome (Groschel et al. 2016; Kato et al. 2017). Several clinical trials, aimed to evaluate the impact of immunotherapy (pembrolizumab, nivolumab or Ipilimumab) on CUP survival, are currently recruiting patients (NCT03391973, NCT03752333, NCT04131621, NCT03396471, NCT02721732).

Moreover, the impact on OS and PFS of novel targeted therapy agents and immunotherapy on unfavorable CUP patients compared to chemotherapy may be highlighted from the results of the ongoing global phase II clinical trial of La Roche-Foundation Medicine (NCT03498521).

2 METHODS

2.1 Patient Sample

A total number of 159 FFPE samples from 150 patients were collected for this study.

Patients were diagnosed and treated at IRCCS Azienda Ospedaliero Universitaria di Bologna, Italy (N = 84), at the University Hospital of Ferrara, Italy (N = 52), or at the Medical University of Graz, Austria (N = 14).

The first cohort consists of patients with tumours with a clearly recognized primary site (N = 104 patients, N = 106 samples) and patients with cancer of unknown or uncertain origin (CUPs, N = 46 patients, N = 53 samples), as summarized in Table 4.

Primary tumours included samples obtained from:

1. Lung
    a. LUAD, adenocarcinoma, N = 6
    b. LUSC, squamous cell carcinoma, N = 3
2. Pancreas
    a. PAAD, exocrine adenocarcinoma, N = 5
3. Ovary
    a. OV, ovarian serous carcinoma, N = 6
4. Liver and biliary tract
    a. LIHC, hepatocellular carcinoma, N = 6
    b. CHOL, cholangiocarcinoma, N = 6
5. Kidney
    a. KICA, kidney renal clear cell carcinoma KIRC, N = 5 and kidney renal papillary cell carcinoma or KIRP, N = 3
6. Colorectum
    a. CRC, adenocarcinoma, N = 7
7. Testis
    a. TGSC, germ cell seminomatous carcinoma, N = 4
8. Endometrium
    a. UCEC, adenocarcinoma, N = 5
9. Stomach
    a. STAD, adenocarcinoma, N = 5
10. Bladder

a. BLCA, transitional cell carcinoma, N = 4

11. Breast
    a. LBC, luminal nonspecial type and lobular breast carcinoma, N = 5
    b. TNBC, triple-negative breast cancer, N = 3
12. Prostate
    a. PRAD, adenocarcinoma, N = 5
13. Melanoma
    a. SKCM, melanoma of skin, N = 7
14. Head and neck
    a. HNSC, squamous cell carcinoma, N = 6
15. Gastrointestinal neuroendocrine carcinoma
    a. GI-NET, N = 5.

Metastases of known origin were a total of 10 and derived from lung, melanoma, stomach, prostate, head and neck, kidney, colon, breast, pancreas, and endometrium.

A total number of 53 CUP samples were included in this study, in particular 43 retrospective and 10 prospective cases. Moreover, from five retrospective CUP patients we obtained metastatic biopsies collected from multiple sites for independent analysis.

For each sample, a full IHC panel was assessed at the time of diagnosis and the outcome was recorded. However, it's important to underline that our collection of CUP samples is heterogeneous since it derives from patients that received the diagnosis in different time; specifically, 14 of them (26%) received the diagnosis of CUP between 2005 and 2009, 26 between 2010 and 2014 (49%), and 13 between 2015 and 2019 (25%).

For each sample, 10 µm thick tissue sections (N=2-5) were obtained. The last section was stained with hematoxylin-eosin (H&E) and examined by an expert pathologist to select the tumour area, which was grossly dissected before RNA extraction. Tumour cell fraction was evaluated to select samples with at least 30% cellularity. Table 5.

			Primaries		Metastases		CUPs	
Characteristics			**n**	***%***	**n**	***%***	**n**	***%***
Patients		n = 151	92		10		49	
Prospective							10	*20*
Retrospective			92		10		39	*80*
Samples		n = 156	96		10		50	
Sex								
	Male	70	45	*49*	3	*30*	22	*45*
	Female	60	29	*32*	5	*50*	26	*53*
	ND	21	18	*20*	2	*20*	1	*2*
Age, years								
	Median		66		71		67	
	Range		44-85		60-86		42-87	
	ND		62		4		2	
Primary tumor classes:								
	Bladder (BLCA)		4					
	Breast (LBC)		5					
	breast (TNBC)		3					
	Cholangiocarcinoma (CHOL)		6					
	Colon (CRC)		7					
	Endometrium (UCEC)		5					
	GI neuroendocrine (GI-NET)		5					
	HeadNeck (HNSC)		6					
	Kidney (KRIC)		5					
	Kidney (KRIP)		3					
	Liver HCC (LIHC)		6					
	Lung -adenocarcinoma (LUAD)		6					
	Lung -squamous (LUSC)		3					
	Melanoma (SKCM)		7					
	Ovary (OV)		6					
	Pancreas (PAAD)		5					
	Prostate (PRAD)		5					
	Stomach (STAD)		5					
	Testis (TGSC)		4					
Metastatic sites								
	bone						1	
	bone marrow						1	
	brain				1		2	
	breast						3	
	colon				1		1	
	duodeno						1	
	kidney						1	
	liver				4		12	
	lung				1			

lung		2
lymph node		12
ND		4
ovary		1
pericardium	1	
pleura		5
prostate		2
skin	1	
soft tissues		2
stomach	1	

Table 4. Summary of samples and patients enrolled in the study

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	Late identification of the primary site	Molecular prediction
CB002	F		65	50	Retrospective	Liver	Adenocarcinoma	ND	POS	NEG	CA125+, chromogranin−, ER−, GATA3−	ND	ND		LBC
CB003	F		81	60	Retrospective	Lymph node	Carcinoma	POS	ND	ND	S100−, CD10−, TTF1−, ER−, PR−, HMB45−, GATA3−	ND	ND		HNSC
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		GI-NET
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

CB013	F		71	70	Retrospective	Lymph node	Mucinous adenocarcinoma	POS	NEG/POS	NEG	ER−, PR−, TTF1−, CDX2−	Gastrointestinal	ND		STAD-CRC
CB014	F		47	50	Retrospective	Lymph node	Papillary adenocarcinoma	POS/NEG	POS/NEG	NEG	vimentin+/−, actin−, CK14+, ER−, PR−, HER2−, TTF1−, thyroglobulin−, WT1−, GATA3+, P40−, PAX8−	TN breast or thyroid	ND		TNBC
CB033	F		77	60	Prospective	Lymph node	Carcinoma	POS	POS	NEG	CK 5−6+, CK14+, GATA3+/−, ER−, PR−, AR, HER2−, P63+, P40+, synaptophysin−, PAX8−, napsin A− WT1−, TTF1−	Sudoriparous gland	Breast		LBC
CB053	F		60	75	Prospective	Liver	Adenocarcinoma	ND	POS	NEG	TTF1+, ALK−, KRAS G34T (pyrosequencing)	Lung	ND		STAD-CRC
CB054	F		59	70	Prospective	Lung	Carcinoma	POS	ND	ND	GATA3+, ER−, PR−, HER2−, Ki67 96%, p40-, p63−, TTF1−, CDX2−	Breast	ND	LBC	LBC
CB055	F		49	50	Prospective	Lymph node	Adenocarcinoma	POS	POS	POS	CDX2+, chromogranin−,	Gastrointestinal	ND		CHOL

											synaptophysin−, CD56−, HER−2−, MSI−				
CB061	M		60	70	Retrospective	Kidney	Adenocarcinoma	POS	POS	ND	AR+, CD10−, OCT4−, PSA−, RCC−, TTF1−, GATA3−, NKX34.1−	Breast or kidney	ND		LBC
CB062	M		87	80	Prospective	Seminal vesicle	Carcinoma	POS	NEG	NEG	CDX2+, SMA−, CD34−, desmin−, HER2−, Ki67:50%, MART1−, MSI−, NKX3.1−, S100−	Gastrointestinal	ND		CHOL/LBC/KICA
CB064	M		58	65	Prospective	Liver	Squamous carcinoma	ND	NEG	NEG	P63+, CK14+, chromogranin−, synaptophysin−, CD45−, S100−	ND	ND		TNBC
CB071	F		64	60	Prospective	Liver	Carcinoma	ND	POS	NEG	CEA+/−, GATA3−, ER−, PR−/+, TTF1−, PAX8−	Biliary duct	ND	CHOL	CHOL
CB090	F		63	30	Prospective	Duodeno	Adenocarcinoma	ND	POS	NEG	CDX2+/−, MUC 1+, MUC 2−, MUC 5AC+, MUC 6+/−, HER2−, MSI−,	Gastrointestinal	ND		STAD-CRC

											synaptophysin−, TTF1−				
CB095	F		70	80	Prospective	Soft tissue	Squamous carcinoma	ND	POS	NEG	CA125−, CA15.3−, CDX2+/−, desmin−, ER−, GATA3−, MUC1+, P40−, PAX−8−, PD−L1 70%, ER−, S100−, synaptophysin−, TTF1−	ND	ND		BLCA
CB097	F		68	90	Retrospective	Liver	Neuroendocrine	ND	ND	ND	chromogranin +, synaptophysin +, CDX2−, gastrin−, glucagon−, insulin−, TTF1−	Gastrointestinal NET	Liver		PAAD
CB098	F	B01	42	80	Retrospective	Lymph node	Squamous cell carcinoma	POS	POS	POS	CD10−, CEA+, GATA3−, HER2−, Ki67 90%, Mammaglobin−, ER−, PR−, AR−, HER2−, BRAF V600E−	Breast	Breast	TNBC	TNBC
CB100	F	B02	42	40	Retrospective	Lymph node	Squamous cell carcinoma	POS	POS	POS	CD10−, CEA+, GATA3−, HER2−, Ki67	Breast	Breast	TNBC	LUAD/TNBC

											90%, Mammaglobin−, TTF1−, ER−, PR−, AR−, HER2−, BRAF V600E−, EBV−(ISH)				
CB101	F	B03	42	80	Retrospective	Breast	Adenocarcinoma	POS	POS	POS	CD10−, CEA+, GATA3−, HER2−, Ki67 70%, Mammaglobin−, Cathepsin K−, ER−, PR−, AR−, HER2−, BRAF V600E−, EBV−(ISH)	Breast	Breast	TNBC	TNBC
CB102	F	B04	42	85	Retrospective	Breast	Adenocarcinoma	POS	POS	POS	CD10−, CEA+, GATA3−, HER2−, Ki67 60%, Mammaglobin−, Cathepsin K−, ER−, PR−, AR−, HER2−, BRAF V600E−, EBV−(ISH)	Breast	Breast	TNBC	LBC
CB103	M		81	50	Retrospective	Lymph node	Large cell carcinoma	POS	POS	POS	CK14+, MUC1+, TTF1−,	Lung	ND		LUAD

											CD10−, CD117−, CD56−, Ki67 50%, EBV−(ISH)				
CB104	M		43	90	Retrospective	Prostate	Squamous cell carcinoma	POS	ND	NEG	PSA−, P40+	Prostate or bladder	Bladder		LUSC
CB105	M	E01	61	40	Retrospective	Liver	Adenocarcinoma	ND	POS	NEG	CDX2−, MUC1+, MUC2+, TTF1−, HER2−	Gastrointestinal	Gastrointestinal		STAD-CRC
CB106	M	E02	61	70	Retrospective	Thyroid	Neuroendocrine	ND	POS	NEG	CD56−, CDX2−, chromogranin+, MUC1+, MUC2+, synaptophysin−, TTF1−	Gastrointestinal	Gastrointestinal		GI-NET, STAD-CRC
CB108	M	F01	74	80	Retrospective	Bone	Adenocarcinoma	POS	POS	NEG	PSA−, TTF1−, CD34−	Upper gastrointestinal tract	Upper gastrointestinal tract		LBC/ STAD-CRC
CB109	M	F02	74	65	Retrospective	Dermis	Undifferentiated carcinoma	POS	POS	NEG	CD31−, CD34−, CK 5/6−, Factor VIII−, LCA−, PSA−, S100−, SOX9−, TTF1−, vimentin−	Upper gastrointestinal tract	ND		STAD-CRC
CB110	F		57	70	Retrospective	Brain	Adenocarcinoma	POS	POS	NEG	BRAF V600E−, TTF1−	Mullerian	Lung		OV
CB112	M		79	35	Retrospective	Colon	Adenocarcinoma	POS	POS	NEG	CDX2−, calretinin−, CDX2−, CEA−,	ND	ND		LUAD

											PDPN−, TTF1−				
CB115	F		86	50	Retrospective	Muscle	Adenocarcinoma	POS	POS	NEG	TTF1−, CDX2−	Intrahepatic bile duct	ND		STAD−CRC/PAAD
CB116	F		65	60	Retrospective	ND	Adenocarcinoma	ND	POS	NEG	TTF1−, CDX2−	Extrahepatic bile duct	ND		STAD-CRC
CB117	M		69	80	Retrospective	Lymph node	Carcinoma	POS	ND	ND	HMB45−, MART1−, S100−, CD10−	ND	ND		LUAD
CB118	M		61	60	Retrospective	Lymph node	Adenocarcinoma	ND	POS	NEG	CDX2−, TTF1−	Pancreas	Lung		LUAD
CB119	M	Q01	66	50	Retrospective	Lymph node	Adenocarcinoma	POS	ND	ND	CK 5/6−, Ki67 50%, MUC1+, PSA−	Prostate or bladder	ND		HNSC
CB120	M	Q02	66	50	Retrospective	Lymph node	Adenocarcinoma	POS	ND	ND	CK 5/6−, Ki67 50%, MUC1+, PSA−	Prostate or bladder	ND		HNSC/LUAD
CB121	M	R01	69	30	Retrospective	Bone	Adenocarcinoma	POS	NEG	POS	BRAF V600E−, CDX2+, PSA−, TTF1−	Gastrointestinal	Bile duct		CHOL
CB122	M	R02	69	65	Retrospective	Liver	Adenocarcinoma	POS	NEG	POS	BRAF V600E−, CDX2+, PSA−, TTF1−	Gastrointestinal	Bile duct		CHOL/PAAD
CB125	M		64	60	Prospective	Cerebellum	Mucinous adenocarcinoma	ND	NEG	ND	PDL1−, TTF1−	ND	Gastrointestinal or lung		STAD-CRC
PF005	F		75	ND	Retrospective	Lung	Poorly differentiated adenocarcinoma	ND	ND	ND	ND	ND	ND		LBC

PF006	F		81	ND	Retrospective	Brain	Clear cell carcinoma	ND	ND	ND	ND	Kidney	Lung		LUAD
PF007	M		53	ND	Retrospective	Liver	Adenocarcinoma	ND	ND	ND	ND	Gastrointestinal	ND		STAD-CRC
PF011	M		75	ND	Retrospective	Lung	Carcinoma with a transitional/squamous and glandular differentiation	ND	ND	ND	TTF1−	ND	ND		LUSC
PF013	F		71	ND	Retrospective	Liver	Poorly differentiated adenocarcinoma	ND	ND	ND	ND	ND	Esophagus		CHOL
PF017	M		72	ND	Retrospective	Liver	Adenocarcinoma	ND	ND	ND	ND	ND	Gallbladder		PAAD
PF018	F		79	ND	Retrospective	Liver	Adenocarcinoma	ND	ND	ND	ND	Pancreas	Small intestine		STAD-CRC
PF019	M		74	ND	Retrospective	Liver	Adenocarcinoma	ND	ND	ND	ND	Pancreas	Pancreas		PAAD/LBC
PF020	M		72	ND	Retrospective	Pleura	Adenocarcinoma	ND	ND	ND	ND	ND	Lung		PAAD
PF021	F		79	ND	Retrospective	Pleura	Adenocarcinoma	ND	POS	ND	TTF1+, CEA+, ER−, PR−	Lung	NA		PAAD
PF022	M		73	ND	Retrospective	Pleura	Adenocarcinoma	ND	ND	ND	CEA+, TTF1−	ND	Gastrointestinal		PAAD
PF024	M		77	ND	Retrospective	Pleura	Adenocarcinoma	ND	ND	ND	TTF1+	Lung	NA		LUSC
PF025	M		50	ND	Retrospective	Pleura	Adenocarcinoma	ND	ND	ND	ND	ND	Lung		LIHC
PF059	F		73	ND	Retrospective	Lymph node	Poorly differentiated carcinoma	POS	POS/ NEG	NEG	TTF1 −, P63−, CD45 −, CDX2 −, MART1−, S100 −, ER+/−, PR −,	Breast	ND		LUAD

											HER2− , Ki67 85%, EBV−(ISH)				
PF080	F		52	ND	Retrospective	Lymph node	Adenocarcinoma	ND	POS	POS	CDX2+, ER+, PR+, CA125−, CD10−, CEA+, chromogranin−, NSE−, TTF1−, vimentin−, WT1−	Breast or genital system	ND		PAAD

Table 5. Prediction outcome in cancer of unknown primary site.

2.2 RNA extraction and cDNA conversion

Total RNA, including microRNAs, was isolated from the tumour FFPE sections using miRNeasy FFPE kit (Qiagen, Hilden, Germany, Cat No. 217504; miRNeasy FFPE Handbook Qiagen, HB-0374-005), following the protocol "Purification of Total RNA, Including miRNA, from FFPE Tissue Sections" in Qiagen miRNeasy FFPE Handbook (v. January 2020) and the Appendix A protocol: "Deparaffinization using xylene, limonene or CitriSolv for deparaffinization".

RNA was eluted in 20–30 µL of nuclease-free water and frozen at −80 °C.

RNA yield and quality were assessed with NanoGenius Spectrophotometer (ONDA Spectrophotometer, Giorgio Bormac s.r.l., Carpi, Italy).

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 µL reaction mix was prepared for each sample mixing:

- 2 µL of 5× reaction buffer
- 4.5 µL of nuclease-free water
- 1 µL of enzyme mix
- 0.5 µL of UniSp6 RNA spike-in
- 2 µL of diluted RNA (for 10 ng of total RNA).

Resulting cDNA was stored in LoBind DNA Eppendorf tubes (Eppendorf, Hamburg, Germany, 0030108051) at −20 °C.

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 prior to digital droplet PCR (ddPCR) analysis.

2.3 MicroRNA selection

miRNA signature for tumour primary site prediction were selected following two published signatures (Ferracin et al. 2011; Rosenfeld et al. 2008) and implemented with 10 additional miRNAs (miR-661, miR-649, miR-24-3p, miR-16-5p, miR-320a, miR-224-5p, miR-423-5p, miR-25-3p, miR-331-3p, and miR-103a-3p).

The additional miRNAs we decided to include in the panel were selected with the aim to test this tool on novel tumor classes or histotypes (CHOL, TGSC, TNBC, and GI-NET), not included in the two previously mentioned studies. Table 6.

microRNA	Plate well	Target sequence	Previous nomenclature	Qiagen LNA assay	Reference
hsa-let-7e-5p	A01	UGAGGUAGGAGGUUGUAUAGUU	hsa-let-7e	YP00205711	(Rosenfeld et al. 2008)
hsa-let-7i-5p	A02	UGAGGUAGUAGUUUGUGCUGUU	hsa-let-7i	YP00204394	(Rosenfeld et al. 2008)
hsa-miR-103a-3p	H12	AGCAGCAUUGUACAGGGCUAUGA	hsa-miR-103-20	YP00204063	(Ferracin and Negrini 2015; Ferracin, Veronese, and Negrini 2010)
hsa-miR-106b-5p	A06	UAAAGUGCUGACAGUGCAGAU	hsa-miR-106b	YP00205884	(Rosenfeld et al. 2008)
hsa-miR-10a-3p	A04	CAAAUUCGUAUCUAGGGGAAUA	hsa-miR-10a*	YP00205688	(Ferracin et al. 2011)
hsa-miR-10a-5p	A03	UACCCUGUAGAUCCGAAUUUGUG	hsa-miR-10a	YP00204778	(Ferracin et al. 2011)
hsa-miR-10b-5p	A05	UACCCUGUAGAACCGAAUUUGUG	hsa-miR-10b	YP00205637	(Rosenfeld et al. 2008)
hsa-miR-122-5p	A07	UGGAGUGUGACAAUGGUGUUUG	hsa-miR-122	YP00205664	(Ferracin et al. 2011)
hsa-miR-124-3p	A08	UAAGGCACGCGGUGAAUGCCAA	hsa-miR-124	YP00206026	(Rosenfeld et al. 2008)
hsa-miR-126-5p	A09	CAUUAUUACUUUUGGUACGCG	hsa-miR-126*	YP00206010	(Ferracin et al. 2011)
hsa-miR-130a-3p	A10	CAGUGCAAUGUUAAAAGGGCAU	hsa-miR-130a	YP00204658	(Rosenfeld et al. 2008)
hsa-miR-135b-5p	A11	UAUGGCUUUUCAUUCCUAUGUGA	hsa-miR-135b	YP00204130	(Ferracin et al. 2011)
hsa-miR-138-5p	B01	AGCUGGUGUUGUGAAUCAGGCCG	hsa-miR-138	YP00206078	(Rosenfeld et al. 2008)
hsa-miR-141-3p	B02	UAACACUGUCUGGUAAAGAUGG	hsa-miR-141	YP00204504	(Ferracin et al. 2011)
hsa-miR-142-3p	B03	UGUAGUGUUUCCUACUUUAUGGA	hsa-miR-142-3p	YP00204291	(Rosenfeld et al. 2008)
hsa-miR-145-5p	B04	GUCCAGUUUUCCCAGGAAUCCCU	hsa-miR-145	YP00204483	(Ferracin et al. 2011)
hsa-miR-146a-5p	B05	UGAGAACUGAAUUCCAUGGGUU	hsa-miR-146a	YP00204688	(Ferracin et al. 2011)
hsa-miR-148b-3p	B06	UCAGUGCAUCACAGAACUUUGU	hsa-miR-148b	YP00203906	(Rosenfeld et al. 2008)
hsa-miR-149-5p	B07	UCUGGCUCCGUGUCUUCACUCCC	hsa-miR-149	YP00204321	(Ferracin et al. 2011)
hsa-miR-152-3p	B08	UCAGUGCAUGACAGAACUUGG	hsa-miR-152	YP00204294	(Rosenfeld et al. 2008)
hsa-miR-16-5p	H06	UAGCAGCACGUAAAUAUUGGCG	hsa-miR-16	YP00205702	(Ferracin and Negrini 2015; Ferracin, Veronese, and Negrini 2010)
hsa-miR-181a-2-3p	B10	ACCACUGACCGUUGACUGUACC	hsa-miR-181a-2*	YP00204142	(Ferracin et al. 2011)
hsa-miR-181a-5p	B09	AACAUUCAACGCUGUCGGUGAGU	hsa-miR-181a	YP00206081	(Rosenfeld et al. 2008)
hsa-miR-181b-5p	B11	AACAUUCAUUGCUGUCGGUGGGU	hsa-miR-181b	YP00204530	(Rosenfeld et al. 2008)
hsa-miR-182-5p	C01	UUUGGCAAUGGUAGAACUCACACU	hsa-miR-182	YP00206070	(Ferracin et al. 2011)

hsa-miR-183-5p	C02	UAUGGCACUGGUAGAAUUCACU	hsa-miR-183	YP00206030	(Ferracin et al. 2011)
hsa-miR-187-3p	C03	UCGUGUCUUGUGUUGCAGCCGG	hsa-miR-187	YP00204018	(Rosenfeld et al. 2008)
hsa-miR-187-5p	C04	GGCUACAACACAGGACCCGGGC	hsa-miR-187*	YP00205920	(Ferracin et al. 2011)
hsa-miR-192-5p	C05	CUGACCUAUGAAUUGACAGCC	hsa-miR-192	YP00204099	(Ferracin et al. 2011)
hsa-miR-193a-3p	C06	AACUGGCCUACAAAGUCCCAGU	hsa-miR-193a-3p	YP00204591	(Ferracin et al. 2011)
hsa-miR-193b-3p	C07	AACUGGCCCUCAAAGUCCCGCU	hsa-miR-193b	YP00204226	(Rosenfeld et al. 2008)
hsa-miR-194-3p	C09	CCAGUGGGGCUGCUGUUAUCUG	hsa-miR-194*	YP00204204	(Ferracin et al. 2011)
hsa-miR-194-5p	C08	UGUAACAGCAACUCCAUGUGGA	hsa-miR-194	YP00204080	(Ferracin et al. 2011)
hsa-miR-196a-5p	C10	UAGGUAGUUUCAUGUUGUUGGG	hsa-miR-196a	YP00204386	(Rosenfeld et al. 2008)
hsa-miR-19b-3p	C11	UGUGCAAAUCCAUGCAAAACUGA	hsa-miR-19b	YP00204450	(Rosenfeld et al. 2008)
hsa-miR-200a-3p	D01	UAACACUGUCUGGUAACGAUGU	hsa-miR-200a	YP00204707	(Ferracin et al. 2011)
hsa-miR-200a-5p	D02	CAUCUUACCGGACAGUGCUGGA	hsa-miR-200a*	YP00206063	(Ferracin et al. 2011)
hsa-miR-200b-3p	D03	UAAUACUGCCUGGUAAUGAUGA	hsa-miR-200b	YP00206071	(Ferracin et al. 2011)
hsa-miR-200b-5p	D04	CAUCUUACUGGGCAGCAUUGGA	hsa-miR-200b*	YP00204144	(Ferracin et al. 2011)
hsa-miR-200c-3p	D05	UAAUACUGCCGGGUAAUGAUGGA	hsa-miR-200c	YP00204482	(Ferracin et al. 2011)
hsa-miR-204-5p	D06	UUCCCUUUGUCAUCCUAUGCCU	hsa-miR-204	YP00206072	(Ferracin et al. 2011)
hsa-miR-205-5p	D07	UCCUUCAUUCCACCGGAGUCUG	hsa-miR-205	YP00204487	(Ferracin et al. 2011)
hsa-miR-210-3p	D08	CUGUGCGUGUGACAGCGGCUGA	hsa-miR-210	YP00204333	(Ferracin et al. 2011)
hsa-miR-211-5p	D09	UUCCCUUUGUCAUCCUUCGCCU	hsa-miR-211	YP00204009	(Ferracin et al. 2011)
hsa-miR-214-3p	D10	ACAGCAGGCACAGACAGGCAGU	hsa-miR-214	YP00204510	(Rosenfeld et al. 2008)
hsa-miR-215-5p	D11	AUGACCUAUGAAUUGACAGAC	hsa-miR-215	YP00204598	(Ferracin et al. 2011)
hsa-miR-21-5p	H05	UAGCUUAUCAGACUGAUGUUGA	hsa-miR-21	YP00204230	(Rosenfeld et al. 2008)
hsa-miR-224-5p	H08	UCAAGUCACUAGUGGUUCCGUUUAG	hsa-miR-224-5p	YP00204641	(Ferracin and Negrini 2015; Ferracin, Veronese, and Negrini 2010)
hsa-miR-24-3p	H04	UGGCUCAGUUCAGCAGGAACAG	hsa-miR-24	YP00204260	(Ferracin and Negrini 2015; Ferracin, Veronese, and Negrini 2010)
hsa-miR-25-3p	H10	CAUUGCACUUGUCUCGGUCUGA	hsa-miR-25	YP00204361	(Ferracin and Negrini 2015; Ferracin, Veronese, and Negrini 2010)
hsa-miR-27b-3p	E01	UUCACAGUGGCUAAGUUCUGC	hsa-miR-27b	YP00205915	(Rosenfeld et al. 2008)
hsa-miR-29a-3p	E02	UAGCACCAUCUGAAAUCGGUUA	hsa-miR-29a	YP00204698	(Rosenfeld et al. 2008)

hsa-miR-29b-3p	E03	UAGCACCAUUUGAAAUCAGUGUU	hsa-miR-29b	YP00204679	(Rosenfeld et al. 2008)
hsa-miR-29c-3p	E04	UAGCACCAUUUGAAAUCGGUUA	hsa-miR-29c	YP00204729	(Rosenfeld et al. 2008)
hsa-miR-30a-3p	E06	CUUUCAGUCGGAUGUUUGCAGC	hsa-miR-30a	YP00204457	(Ferracin et al. 2011)
hsa-miR-30a-5p	E05	UGUAAACAUCCUCGACUGGAAG	hsa-miR-30a*	YP00205695	(Ferracin et al. 2011)
hsa-miR-30c-5p	E07	UGUAAACAUCCUACACUCUCAGC	hsa-miR-30c	YP00204783	(Ferracin et al. 2011)
hsa-miR-31-3p	E09	UGCUAUGCCAACAUAUUGCCAU	hsa-miR-31*	YP00204079	(Ferracin et al. 2011)
hsa-miR-31-5p	E08	AGGCAAGAUGCUGGCAUAGCU	hsa-miR-31	YP00204236	(Ferracin et al. 2011)
hsa-miR-320a	H07	AAAAGCUGGGUUGAGAGGGCGA	hsa-miR-320a	YP00206042	(Ferracin and Negrini 2015; Ferracin, Veronese, and Negrini 2010)
hsa-miR-331-3p	H11	GCCCCUGGGCCUAUCCUAGAA	hsa-miR-331-3p	YP00206046	(Ferracin and Negrini 2015; Ferracin, Veronese, and Negrini 2010)
hsa-miR-340-3p	E10	UCCGUCUCAGUUACUUUAUAGC	hsa-miR-340*	YP00204250	(Ferracin et al. 2011)
hsa-miR-342-3p	E11	UCUCACACAGAAAUCGCACCCGU	hsa-miR-342-3p	YP00205625	(Ferracin et al. 2011)
hsa-miR-345-5p	E12	GCUGACUCCUAGUCCAGGGCUC	hsa-miR-345	YP00206006	(Rosenfeld et al. 2008)
hsa-miR-34a-5p	F01	UGGCAGUGUCUUAGCUGGUUGU	hsa-miR-34a	YP00204486	(Rosenfeld et al. 2008)
hsa-miR-34b-3p	F02	CAAUCACUAACUCCACUGCCAU	hsa-miR-34b	YP00204005	(Rosenfeld et al. 2008)
hsa-miR-361-5p	F03	UUAUCAGAAUCUCCAGGGGUAC	hsa-miR-361-5p	YP00206054	(Ferracin et al. 2011)
hsa-miR-363-3p	F04	AAUUGCACGGUAUCCAUCUGUA	hsa-miR-363	YP00204726	(Ferracin et al. 2011)
hsa-miR-372-3p	F05	AAAGUGCUGCGACAUUUGAGCGU	hsa-miR-372	YP00204137	(Rosenfeld et al. 2008)
hsa-miR-373-3p	F06	GAAGUGCUUCGAUUUUGGGGUGU	hsa-miR-373	YP00204604	(Rosenfeld et al. 2008)
hsa-miR-375-3p	F07	UUUGUUCGUUCGGCUCGCGUGA	hsa-miR-375	YP00204362	(Ferracin et al. 2011)
hsa-miR-382-5p	F08	GAAGUUGUUCGUGGUGGAUUCG	hsa-miR-382	YP00204169	(Rosenfeld et al. 2008)
hsa-miR-423-5p	H09	UGAGGGGCAGAGAGCGAGACUUU	hsa-miR-423-5p	YP00205624	(Ferracin and Negrini 2015; Ferracin, Veronese, and Negrini 2010)
hsa-miR-485-5p	F09	AGAGGCUGGCCGUGAUGAAUUC	hsa-miR-485-5p	YP02112548	(Ferracin et al. 2011)
hsa-miR-506-3p	F10	UAAGGCACCCUUCUGAGUAGA	hsa-miR-506	YP00204539	(Ferracin et al. 2011)
hsa-miR-508-3p	F11	UGAUUGUAGCCUUUUGGAGUAGA	hsa-miR-508-3p	YP00204480	(Ferracin et al. 2011)
hsa-miR-509-3p	F12	UGAUUGGUACGUCUGUGGGUAG	hsa-miR-509-3p	YP00204458	(Ferracin et al. 2011)
hsa-miR-510-5p	G01	UACUCAGGAGAGUGGCAAUCAC	hsa-miR-510	YP00204349	(Ferracin et al. 2011)
hsa-miR-514a-3p	G02	AUUGACACUUCUGUGAGUAGA	hsa-miR-514	YP00205931	(Ferracin et al. 2011)

hsa-miR-552-3p	G03	AACAGGUGACUGGUUAGACAA	hsa-miR-552	YP00206032	(Ferracin et al. 2011)
hsa-miR-649	G04	AAACCUGUGUUGUUCAAGAGUC	hsa-miR-649	YP00204146	(Ferracin et al. 2011)
hsa-miR-650	G05	AGGAGGCAGCGCUCUCAGGAC	hsa-miR-650	YP00204233	(Ferracin et al. 2011)
hsa-miR-661	G06	UGCCUGGGUCUCUGGCCUGCGCGU	hsa-miR-661	YP00204657	(Ferracin et al. 2011)
hsa-miR-873-5p	G07	GCAGGAACUUGUGAGUCUCCU	hsa-miR-873	YP00204175	(Ferracin et al. 2011)
hsa-miR-92a-3p	G09	UAUUGCACUUGUCCCGGCCUGU	hsa-miR-92	YP00204258	(Rosenfeld et al. 2008)
hsa-miR-92b-3p	G10	UAUUGCACUCGUCCCGGCCUCC	hsa-miR-92b	YP00204384	(Rosenfeld et al. 2008)
hsa-miR-9-3p	G08	AUAAAGCUAGAUAACCGAAAGU	hsa-miR-9*	YP00204620	(Rosenfeld et al. 2008)
hsa-miR-96-5p	G11	UUUGGCACUAGCACAUUUUUGCU	hsa-miR-96	YP00204417	(Ferracin et al. 2011)
hsa-miR-99a-5p	G12	AACCCGUAGAUCCGAUCUUGUG	hsa-miR-99a	YP00204521	(Rosenfeld et al. 2008)
SNORD44 (hsa)	H01			YP00203902	
SNORD48 (hsa)	H02			YP00203903	
U6 snRNA (hsa)	H03			YP00203907	
UniSp3 IPC	A12			YP02119288	
UniSp3 IPC	D12			YP02119288	
UniSp6 CP	B12			YP00203954	
NTC	C12				

Table 6. List of miRNA assays in the custom ddPCR plate and miRNAs used by PAMr and LASSO for the site-of-origin prediction

2.5 Droplet digital PCR

Prespotted 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 (Laprovitera et al. 2018).

EvaGreen-based droplet digital PCR was performed as described (Laprovitera et al. 2018). Thermal cycling conditions were: 95 °C for 5 min, then 40 cycles of 95 °C for 30 s and 58 °C for 1 min, and final steps at 4 °C for 5 min, 90 °C for 5 min and a 4 °C infinite hold.

Droplet selection was performed individually for each well using quantasoft software v 1.7 (Bio-Rad, Hercules, CA, USA).

Final miRNA amounts (copies·μL^{-1}) were obtained and normalized on 50th percentile expression using gx v.14.9.1 software (Agilent Technologies, Santa Clara, CA, USA).

None of the candidate reference RNAs included in the plate were used as normalizer due to the higher variability than median expression.

2.6 Tissue-of-origin prediction

Primary tumors (N = 96) were used as training set (Ferracin et al. 2011). Digital droplet PCR data were normalized on the 50th percentile using genespring gx v.14.9.1 software (Agilent Technologies).

Two different methods of prediction have been applied to select the discriminant miRNAs and to predict the tissue of origin, namely PAM and LASSO.

In particular, PAM method uses a shrinkage nearest neighborhood centroid approach in the space of the samples. In the training set, PAM calculates centroids as the standardized gene expression within each class (mean divided by standard deviation). Then, a procedure of shrinkage is applied to move centroid toward zero by a quantity called threshold that is set by the user. This threshold is selected based on the results of cross-validation technique to minimize the error rate. If the shrinkage process reduces to zero the centroid of a gene across all the classes, the gene is not selected for the prediction step. Then, in the prediction step the distance between the expression profile of a new sample with all the class centroid is calculated and the new sample is predicted to belong to the closest one.

Instead, LASSO regression is based on a linear regression model where the objective function is penalized by the sum of the absolute value of the parameters. The dependent variable is the class of the samples, and genes are the covariates of the model. The penalization approach has the effect to shrink the parameters estimate toward zero. If the shrinkage procedure set the parameter to zero, the gene will not be used for the prediction. The magnitude of the penalization is selected using cross-validation technique.

These classifiers were subsequently used to predict known and unknown/uncertain metastases tissue of origin.

Both models assign to every metastatic tumour a probability to be originated from each primary.

Results were compared with the indications of a possible primary site suggested by standard diagnostic immunohistochemistry and clinicopathological assessment.

Subsequently, cluster analysis was performed using the normalized expression (50th percentile) of the 89 miRNAs first in individual patients of the reference set of primary tumours, then averaged levels within each tumour in reference set.

Hierarchical cluster analyses were performed using genespring gx v.14.9.1 software (Agilent Technologies) using complete-linkage rule and Manhattan correlation distance.

Standard deviation on the average expression of each miRNA within each class was also assessed.

2.7 Cell culture

Two patients of our cohort with confirmed diagnosis of cancer of unknown primary (CB055 and CB096) developed refractory ascites, treated with peritoneal drainage.

About 100 ml of ascites from both patients were collected and immediately processed.

Samples were centrifuged at 600 g for 10 minutes, pellet was treated with 10 ml red blood cell lysis buffer Miltenyi Biotec (15 minutes at room temperature).

10 mL of specific culture medium (DMEM/F12 50:50, 2 mM glutamine, 5% Horse serum, hydrocortisone 1 ug/ml, insuline 10 µg/ml, 1X PEN/STREP, EGF 10 ng/ml) was added.

Cells, after centrifugation, were plated in T25 flask with 5 ml of medium in humidified 37 °C/5% CO2 incubator.

Medium was changed every 3–4 days.

After 10 days, tumor cells begun to grow, CB055 as spheroid, CB096 as spheroids.

Organoids and aggregates were propagated in 24 well-plates and long-term cell lines were established (Kodack et al. 2017).

2.8 Patient derived xenograft

PDX studies were run at XenTech in compliance with authorization n. APAFIS#30365-2021012215599431 v1 conferred from the French Ministry of Agriculture and Food. The authorization to use animals in the CERFE facility (Evry-Courcouronne, France) was obtained by The Direction Départementale de la Protection des Populations, Ministère de l'Agriculture et de l'Alimentation, France "Direction of the Veterinarian Services, Ministry of Agriculture and Food, France" (agreement No. D-91-228-107).

An approximate number of 3000-4000 cells or 30-40 organoids were resuspended in 100 µl of culture medium, diluted 1:1 in matrigel and grafted in the interscapular region of NOD/Scid-IL2Rγ-/- (NSG) or NOD/Shi-scid/IL-2Rγnull (NOG) mice. Tumor fragments were sampled from the resected tumor, minced on ice and immediately placed in MACS Tissue Storage Solution (Miltenyi Biotech) and transferred to the animal facility, fragmented and grafted in the interscapular region of Athymic Nude-Foxn1nu mice.

Mice were monitored twice weekly for signs of tumor growth. Tumor growth from first implantation occurred in 135 days for CB055 and 40 days for CB096. Growing tumors were serially transplanted onto recipient mice and the fragments of the tumor harvested for IHC analysis, DNA and RNA extraction. To immortalize each PDX, vials of tumor fragments at different passages were placed in 90% FCS/10% DMSO or glycerol, and stored at -150°C.

2.9 Pathological analysis

Patient sample, cell lines and PDX tumour tissue were fixed in 10% buffered formalin, then paraffin embedded, sections of 4 µm were mounted on positively charged microscope slides. Sections were deparaffinized in xylene and rehydrated in graded alcohol. Antigen enhancement was done by incubating the sections in Concentrated Antigen Retrival Solution Citra Plus (BioGenex #HK080-5K) (1:10) and blocked with Ventana Antibody Diluent with Casein (Roche-Ventana) for 30 minutes.

For immunohistochemistry: Rabbit Monoclonal Primary Antibody Cytokeratin 7 (SP52) (Roche-Ventana) and Rabbit Monoclonal Primary Antibody Cytokeratin 20 (SP33) (Roche-Ventana), were used for patient, cell lines and PDX tumor tissues staining.

For immunofluorescence experiment: Mouse Monoclonal Primary Antibody CD44 (or HCAM) sc-7297 (SantaCruz) 1:100 dilution and Mouse Monoclonal Primary Antibody EpCAM FITC (REA764) (Miltenyi) 1:100 dilution was used for cell lines staining. All the antibodies were incubated overnight at 4°C. AbCAM Goat Anti-Rabbit IgG H&L (DyLight® 594) (ab96885) and Donkey Anti-Mouse IgG H&L (DyLight® 488) (ab96875) labeled secondary antibodies were used at 1:100 dilution and incubated 1h. Nuclei were counterstained with 1 µg/mL of Hoechst 33342 (Life Technologies).

Confocal images were acquired on Nikon A1 confocal laser scanning microscope, equipped with a 60X, 1.4 NA objective and with 405, 488 nm laser lines.

2.10 Genetic analysis

We used a custom 1.2-Mb SureSelect capture bait library (Agilent Technologies, Santa Clara, CA, United States) for the target enrichment of 92 genes, as described (Laprovitera et al. 2018).

Libraries were prepared using 50 ng of gDNA input following SureSelectXT HS/SureSelectXT Low Input Target Enrichment with Pre-Capture Pooling protocol (G9702-90005, v. A0, June 2019, Agilent Technologies) and sequenced on NextSeq 500 (Illumina) platform using High Output 2 × 75-bp flow cells. Variant calling and paired analyses (tumor vs. normal) were performed using SureCall software (v. 4.2), applying a filter for tumor/normal tissue/models at 5% and for ccfDNA at 1%. Variants were annotated using ANNOVAR (Wang, Li, and Hakonarson 2010) and filtered to keep somatic exonic non-synonymous single-nucleotide variants (SNVs), insertions, deletions, multiple nucleotide variants, or long deletions not detected in the normal sample that presented an allele frequency in Non-Finnish European (NFE) population lower than 0.5% (Genome Aggregation Database, GnomAD; Karczewski et al., 2020) and a coverage higher than 100. Bioinformatic pathogenicity prediction, reported in Tables 1 and 2 of the identified variants was performed consulting the prediction score/outcome of 8 prediction models: SIFT (Sort Intolerated From Tolerated; Vaser et al., 2016), Polyphen2 HVAR (Polymorphism Phenotyping v2; Adzhubei et al., 2010), LRT (Likelihood Ratio Test; Chun and Fay, 2009), Mutation Taster (Schwarz et al., 2014), Mutation Assessor (Reva et al., 2011), FATHMM (Functional Analysis Through Hidden Markov Model; Shihab et al., 2013), CADD (Combined Annotation Dependent Depletion; Kircher et al., 2014), and VEST (Variant Effect Scoring Tool; Douville et al., 2016). Variants were considered pathogenic when more than 50% of the above-mentioned predictors indicated it as pathogenic/damaging/deleterious/harmful.

2.11 Fluorescence In-Situ Hybridization (FISH)

FISH analysis was performed on CB055 and CB096 with two Empire Genomics probes (Empire Genomics, Buffalo, NY). FGFR2 gene probe (orange) maps on chr:10q26.13 and the CEP10 probe maps to the centromeric region of Chromosome 10.

Cells were counterstained with 40,6-diamidino-2-phenylindole (DAPI) for nuclear detection. Analysis was performed using Olympus BX53 microscopy equipped with the appropriate filter sets and CytoVision software (Leica Biosystems, Nussloch, Germany).

2.12 Detection of FGFR2 amplification

Total DNA was extracted cell lines and PDXs from CB055 and CB096 using QIAamp DNA Mini Kit (Cat. No. 5130450, Qiagen). Cell free DNA was extracted from 1mL plasma with Maxwell RSC ccfDNA Plasma Kit (Cat No: AS1480, Promega). Tumor DNA was extracted using the QIAamp DNA FFPE Tissue Kit (Cat No: 56404, Qiagen, Hilden, Germany). To quantify the copy number of FGFR2 in all samples, a probe-based droplet digital PCR assay was used. RPP30 was tested as reference gene for diploid copy number. RPP30 probe (dHsaCP2500350, Bio-Rad) was labeled with HEX, and FGFR2 probe (dHsaCB2500320) was labeled with FAM. Droplet digital PCR was performed with the QX200 Droplet Digital PCR system (Bio-Rad, USA) as described in (10). FGFR2 gene copy number was calculated by QuantaSoft Analysis software (Bio-Rad, Hercules, CA, USA) as FGFR2/RPP30 ratio.

2.13 Cell Viability and Drug response assays

Cell viability/proliferation was evaluated by a CellTiter-Glo 96 Aqueous One solution Assay (Promega). Cell death was analyzed by fluorescence microscopy after staining with Hoechst 3342 and Propidium Iodide (PI) (Fumarola et al. 2005). BGJ398 (infigratinib) and trametinib (mekinist) were purchased from Selleckchem (Houston, TX), and dissolved in DMSO. DMSO concentration never exceeded 0.1% (v/v); equal amounts of the solvent were added to control cells. The effect of the drug combination was evaluated using the Bliss interaction model (La Monica et al. 2009).

2.14 In vivo treatments

Treatment efficacy study on each PDX were run as follows.

Tumor fragments of the same passage were transplanted subcutaneously onto 3-24 mice (donor mice). When the tumors reached 700 to 1764 mm^3, donor mice were sacrificed, and tumors were cut into fragments measuring approximately 20 mm^3.

Mice aged 8 to 11 weeks were anaesthetized with 100 mg/kg ketamine hydrochloride and 10 mg/kg xylazine, and one tumor fragment was placed in the interscapular fat pad.

On the day of enrollment in the study mice with tumor volume ranging 60 to 200 mm^3 were randomly attributed to the different groups:

Gr.	N	1st Testing Agent					2nd Testing Agent				
		Agent	Dose mg/kg	Route	Volume ml/kg	Schedule	Agent	Dose mg/kg	Route	Volume ml/kg	Schedule
1	5	Vehicle 1	-	PO	10	qd	Vehicle 2	-	PO	10	qd
2	5	trametinib	0.6*	PO	10	qd	-	-	-	-	-
3	5	-	-	-	-	-	BGJ398	15*	PO	10	qd
4	5	trametinib	0.6	PO	10	qd	BGJ398	15	PO	10	qd

Legend: PO= per os; qd= once daily

From day 12 to day 14 half dose was administered due to body weight loss.

Trametinib was purchased at Carbosynth and suspended in 10% DMSO; 40% PEG300; 5% Tween-80; 45% NaCl 0.9% for administration; BGJ398 (infigratinib) was purchased at MedChem Express and diluted in 50% Acetic acid/acetate buffer / 50% PEG300 for administration. Tumor volume was evaluated by measuring tumor diameters with a caliper, three times a week during the treatment period (from D0 to D27); all animals were weighted, and tumor size measured the same day. During the whole experimental period, animals were monitored every day for physical appearance, behavior, and clinical changes.

2.15 Single Cell Transcriptomic Analysis

The transcriptome of CB055 cell line was characterized at the single-cell level. To this end, we relied on the 10X Single Cell 5' R2-only kit. The resulting libraries were sequenced with the Novaseq 6000 (Illumina) platform using a 2 × 100-bp flow cell, obtaining a total yield of 240 million reads. Gene expression profiles of different cells were quantified by the CellRanger pipeline (version 7.1.0), using as a reference the GENCODE annotation (version 40). The software provided an initial estimate of 23,994 sampled cells. Gene counts were filtered by means of the CellBender method to remove the confounding effects of ambient RNA molecules and random barcode swapping. Around 7% of all droplets were identified by scDblFinder (version 1.14.0) as potential doublet artifacts and were excluded from the rest of the analysis. We also discarded cells characterized by a high proportion of reads from the mitochondrial genome, a signal usually associated with cell membrane rupture and apoptosis. We set our filtering threshold to 50% following the indications of other studies that have

observed very high mitochondrial content in hepatocytes. As a last QC filter, we dropped all cells showing a detectable level of expression for less than 500 genes. The final working set consisted of 1,437 cells.

Gene expression was loaded in the R environment using Seurat (version 4.3.0.1) and normalized with the sctransform method (v2) to account for gene overdispersion. Cells were clustered using the Latent Cellular Analysis method (scLCA package, version 0.0.0.9000), which combines a cosine-similarity measure with a graph-based algorithm to automatically identify distinct cell subpopulations. Annotation of cell types was performed using the Azimuth software (version 0.4.6) and relied on its reference atlas of the human liver.

We applied the MAST software to identify marker genes of cell clusters. This method (version 1.24.0) implements a generalized linear model specifically designed to account for nuisance variation in scRNA-seq data. We filtered differential expression calls by setting an FDR threshold of 10-3, and we also discarded all genes whose log2 fold-change lied below 0.2.

Functional analysis of the resulting gene sets was performed using the online tool DAVID, computing the over-representation probability of marker genes within pathways annotated in Biocarta, KEGG, Reactome and WikiPathways. Ribosomal genes were excluded from this analysis.

3 RESULTS

3.1 miRNA testing with droplet digital PCR

miRNAs are extremely stable in formalin-fixed, paraffin-embedded tissue, therefore FFPE sample are the most commonly available source of tumor material in clinical setting.

We developed custom multi-miRNA expression assay capable of testing the absolute levels of 89 miRNAs compatible with standard diagnostic workup and with the amount of available material.

This multi-miRNA assay is based on absolute miRNA quantification with EvaGreen Dye Droplet Digital PCR technology (Laprovitera et al. 2018).

RNA was extracted from 2 to 5 slices of tumor FFPE blocks, tumor area was identified by experienced pathologists and macrodissected. An amount of 10 ng is sufficient to test all miRNAs in a single experiment.

We obtained the absolute copy number for all miRNAs in our panel in the same experiment, with identical experimental conditions.

We tested 96 primary tumors as reference set for molecular profiling. This set was composed with 16 different tumor types with 19 histological classes, focusing on the most common CUP's sites of origin described in autopsy study (Pentheroudakis, Golfinopoulos, and Pavlidis 2007).

The expression matrix of the primary tumor dataset created 19 different tumors classes: LUAD, LUSC, PAAD, LIHC, CHOL, KIRC, KIRP, STAD, CRC, TGSC, OV, UCEC, BLCA, LBC, TNBC, PRAD, SKCM, GI-NET, and HNSC.

3.2 miRNA expression patterns

Average and normalized expression of the 89-miRNA signature in the nineteen primary tumor types were reported analyzed with cluster analysis in Figure 1 and Supplementary Table 1.

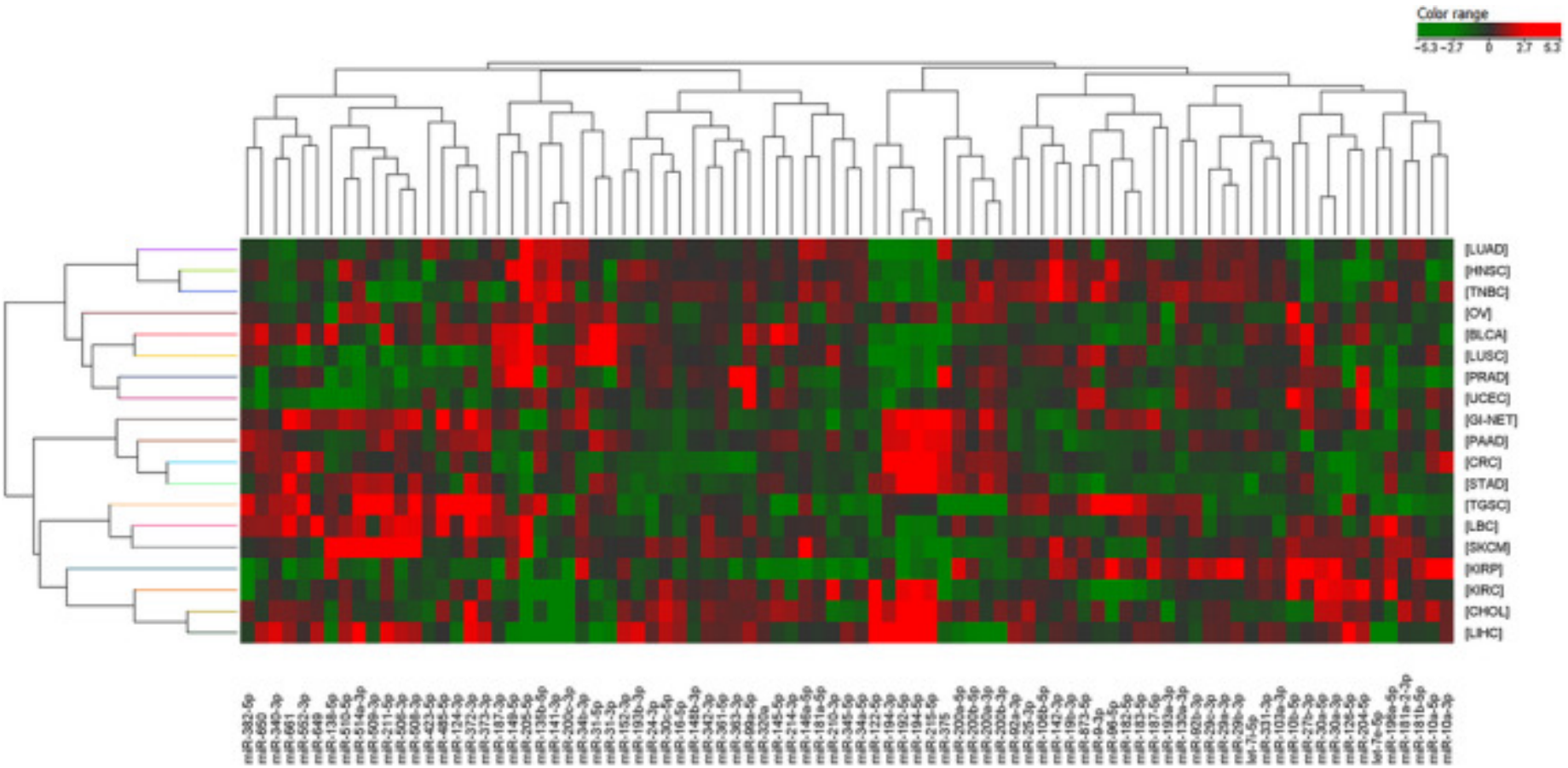

Figure 1. Cluster analysis of primary tumors. Heatmap representing the expression of 89 microRNAs in 19 different classes of primary tumors. Averaged, normalized miRNA levels in each tumor class were used for clustering analysis. Green indicates low expression, and red indicates high expression.

In our cohort each tumor type displays a peculiar pattern of miRNA expression.

Nonetheless, we found some unexpected similarities and divergences among tumor types:

1. STAD and CRC was found to be consistently overlapping and partially intermixed with other gastrointestinal tumors (PAAD and GI-NET), as previously reported (Ferracin et al. 2011; Rosenfeld et al. 2008; Sokilde et al. 2014). Due to we decided to consider them as a single class (STAD-CRC) for molecular prediction.

2. Renal KIRC and KIRP, showing similar miRNA expression, were combined in the tumor class KICA.

3. Female reproductive-system tumours (OV and UCEC) were found to express similar yet distinct miRNA patterns (Ferracin et al. 2011; Rosenfeld et al. 2008; Sokilde et al. 2014).

4. Triple negative breast cancer (TNBC) share a portion of their signatures with lung cancers (LUAD and LUSC), but not with luminal breast cancer. They also showed an unexpected similarity with HNSC. We could speculate a possible etiology associated to human papillomavirus (HPV) infection has been reported in both these tumor types (De Carolis et al. 2019; Otter et al. 2019; Piana et al. 2014).

In conclusion, this signature is reliable in discriminating among 17 different tumour.

3.3 miRNA prediction

We applied the nearest shrunken centroids (NSC) using PAMR (Tibshirani 2009) and the least absolute shrinkage and selection operator (LASSO) predictive models (Tibshirani 1996) developed by Tibshirani's laboratory. 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 2. Overall error rate for both PAMR and LASSO was 33%. However, 11 of the 17 tumor classes (LIHC, LUSC, LBC, KICA, GI-NET, TGSC, STAD-CRC, SKCM, LUAD, UCEC, and PRAD) had error rates much lower with both models (17% for PAMR and 22% for LASSO). Of note, PAMR seems to be considerably more accurate in the prediction of LBC and LUSC compared to LASSO; on the contrary, LASSO seems to be 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 reduced specificity of the miRNA signature for these primaries and cross-prediction (e.g., CHOL and PAAD or TNBC and HNSC) or the smaller sample size of TNBC (n = 3) and BLCA (n = 4).

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

Sample ID	Primary site	PAMR 1st prediction	PAMR 2nd prediction	LASSO 1st prediction	LASSO 2nd prediction
CB048	LBC	LBC	PAAD	LBC	STAD-CRC
CB044	CRC	STAD-CRC	PAAD	STAD-CRC	LIHC
PF088	UCEC	OV	PAAD	PAAD	PRAD
PF086	STAD	STAD-CRC	PAAD	STAD-CRC	PAAD
CB073	HNSC	TNBC	HNSC	TNBC	HNSC
PF095	KIRC	CHOL	LIHC	KICA	PAAD
PF010	LUAD	LBC	LUAD	LUAD	PAAD
CB049	PAAD	STAD-CRC	PAAD	STAD-CRC	PAAD
PF087	PRAD	STAD-CRC	PAAD	STAD-CRC	PAAD
PF079	SKCM	LUAD	SKCM	TNBC	KICA

Then the two models were used to predict the primary site of 53 CUPs. Histopathological and immunohistochemistry of this group is reported in Table 4 and 5.

Tumour site prediction is represented in Figure 2, in which the top two primary sites predicted by both models.

PAMR reached a reliable molecular prediction in 43 out of 53 CUPs (81%). Instead, LASSO predicted 25 out of 53 CUPs (47.2%).

The most probable primary sites, reported in Table 4, were prioritized using the following criteria:

1. The primary site was predicted by at least one predictive model (LASSO or PAMR) with a probability higher than 80%;
2. The primary site was present among the predicted sites in both models, with a probability higher than 30% in at least one prediction.

If the prediction outcome did not fall within these criteria, we reported all the predicted primary sites.

Few cases had more than one tissue of origin. The same primary sites were predicted by both models in 94% of cases.

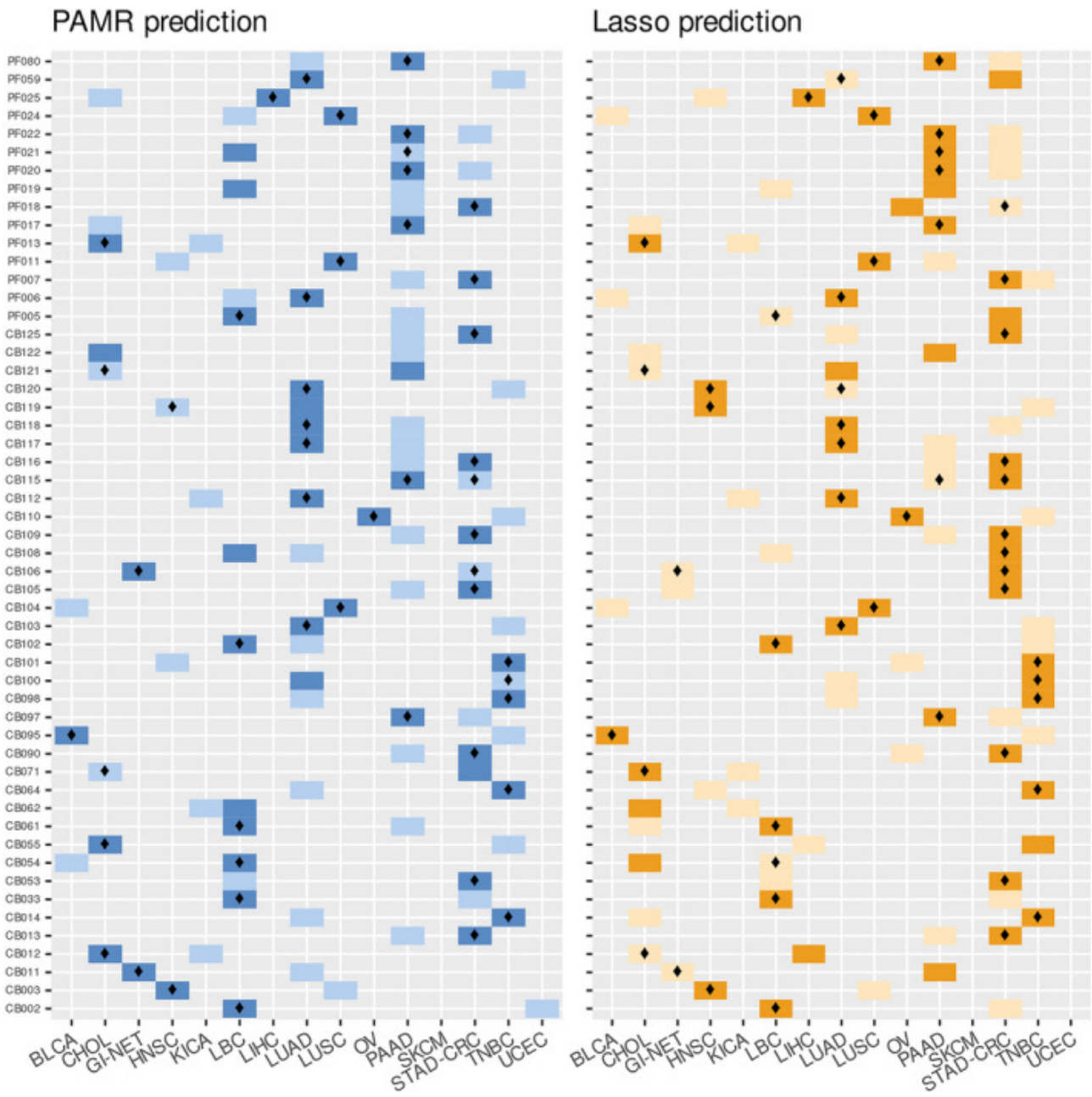

Figure 2. Prediction outcome of cancers of unknown origin using PAMR NSC and LASSO classifiers. For each of the 53 CUP sample (on the y-axis), the two top predicted primary tumors (x-axis) are highlighted. PAMR first and second molecular predictions are reported with dark and light-blue squares, respectively. 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.

We also evaluated the compatibility of this molecular prediction with the clinicopathological information available. Final predictions were found in agreement with the first hypothesis of a primary site in 53% of CUPs in which a hypothesis was made.

In patients in which the primary site was later identified (N = 3, CB071, CB054, and CB098/100/101/102), we observed a complete concordance between the diagnosis and the molecular prediction.

Considering the final predicted sites reported in Table 2, the most common tissues of origin were:

1. STAD-CRC (19%)
2. LBC (15%)
3. PAAD (15%)
4. LUAD (13%)
5. CHOL (11%)
6. LUSC (5%)
7. TNBC (8%)
8. HNSC (5%)
9. TGSC (0%)
10. PRAD (0%)

3.4 CUP models from ascites tumours cells

Two CUP patient developed ascites.

First patient (CB055) was diagnosed with multiple lymph node metastases of poorly differentiated adenocarcinoma. Immunohistochemistry (IHC) staining reported the positivity for keratin 7 (K7) and CDX2, weak positivity for keratin 20 (K20) and negativity for neuroendocrine markers (chromogranin, synaptophysin, CD56).

The second patient (CB096) presented with multiple metastases of poorly differentiated adenocarcinoma with peritoneal carcinosis and multiple lymph node metastases. IHC staining was positive for K7, K20, CDX2, EpCAM and negative for PAX8, p40, GATA3 and calretinin.

Two long-term (more than 10 passages (25)) cell lines were obtained isolating cells from ascites fluid of both patients (CB055 and CB096). One month after seeding, cells growing in suspension were visible as spheroid-like structures. The growth curve of two models was monitored for 10 days using the Incucyte S3 live-cell analysis: 10000 cells of CB096 and CB055 were seeded and after few

hours CB096 cells formed large tumoroids and CB055 cells organized as clusters, showing a doubling time of 5 days for both cell lines.

We also generated two PDX models by injecting ascites tumor cells in the interscapular region of immunocompromised mice. The two models showed different growth rate, with CB096 PDX growing faster than CB055 PDX, and with CUP055 PDX inducing a tumor-intrinsic mild cachexia in the mice.

3.5 Histology and immunohistochemistry from models

Histology of the cell lines and PDXs recapitulated the features of the primary tumor (Fig. 3A-B). To verify whether CUP tumor cells presented a stem-cell like phenotype, the two tumoroid cell lines were tested for CD44 and EPCAM immunoreactivity and the expression of stemness genes CD44, NANOG and OCT4 (Fig. 4. The two cell lines express CD44 on their surface, which is more homogeneous for CB096 (Fig. 4A), and both express stemness genes at high levels if compared to a panel of cancer cell lines (Fig. 4B).

Predictive microRNA-based test of CB055 was predicted to be of biliary tract origin (probability of 93%) and CB096 of gastrointestinal tract origin (probability of 99%).

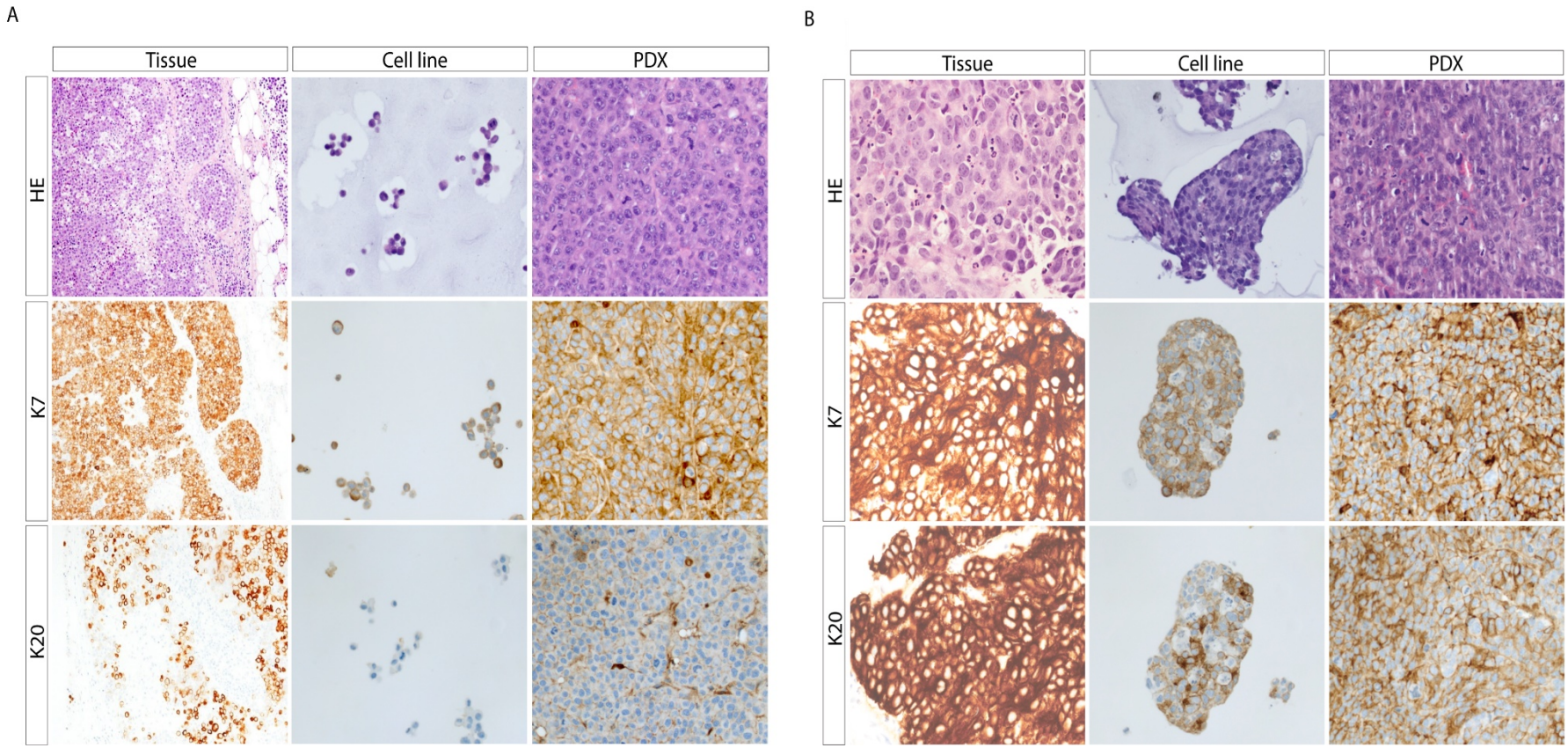

Figure 3. A) Immunophenotypic characterization of tumor tissues, cell lines and PDX of CB055. Hematoxylin-eosin (HE), keratin7 (K7) and keratin20 (K20) IHC staining was reported for tumor tissue, cell lines and PDX in CB055 model (40x). The staining showed that cell line and PDX models recapitulated the histology of the tumor tissue. B) Immunophenotypic characterization of tumor tissues, cell lines and PDX of CB096. Hematoxylin-eosin (HE), keratin7 (K7) and keratin20 (K20) IHC staining was reported for tumor tissue, cell lines and PDX in CB096 model (40x). CB096 cell line grows forming tumoroid structures. The staining showed that cell line and PDX models recapitulated the histology of the tumor tissue.

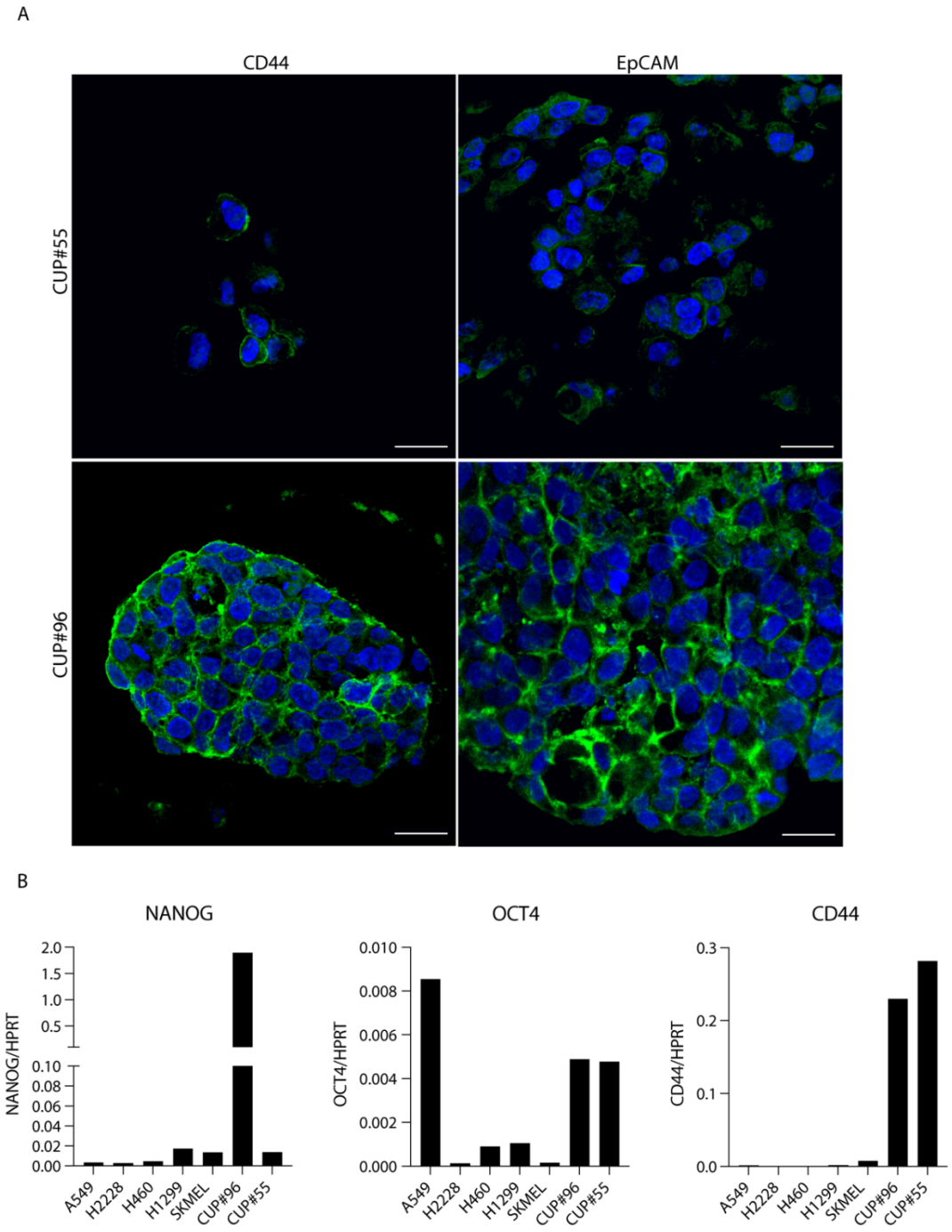

Figure 4. CUP models characterization. A) Immunophenotypic analysis of CUP #55 and #96 cell lines demonstrating the expression of cancer stem cell markers CD44 and EPCAM. Staining: nucleus (DAPI, blue), CD44 and EPCAM (green). B) Gene expression analysis of stemness genes in CUP#55 and CUP#96. Bars represent the ratio between the gene copies and the reference HPRT copies assessed by ddPCR. The expression in CUP cell lines was compared to other cancer cell lines from lung and melanoma origin.

3.7 Genetic characterization of patient and models tumours

The tumor tissue, circulating cell-free DNA (ccfDNA), cell line and PDX were tested for genetic alterations with a CUP-dedicated, 92-gene custom panel using SureSelect Target Enrichment technology (Agilent Technologies) as described (Laprovitera, Salamon, et al. 2021).

Genetic alterations are reported in Table 7 and 8.

The genetic analysis of CB055 harbour 5 genetic alterations in all the analyzed samples (bulk tumor DNA, ccfDNA, cell line and PDX): the insertion in the APC gene (p.T1556fs*3) and the point mutations in ARID1A (p.R1276∗), ERBB3 (p.S128R), KEAP1 (p.R135H) and NTRK1 (p.Q736X). Mutations in ALK, EPHA5, FAT1, KMT2C, MGA, PTPRD (except for p.L970V) and TP53 were detected only in ccfDNA (Table 7). As expected, the variant allele fractions (VAFs) were higher in PDX and cell line, thus reflecting the greater tumor purity and the possible selection of tumor subclones. Low frequency mutations in ZFHX3, RBM10, PTPRT, NF1, MED12, MGA, KDR, KDM5A, GRIN2A, EP300, DOT1L, APC (p. S887R), were identified in the models but not in other samples.

The generic analysis of CB096 was conducted only on ccfDNA and cell line due to the unavailability of residual tumor sample for this patient. Five genetic alterations were detected in both samples (Table8), with higher VAFs in the cell line compared to ccfDNA.

Low-pass Copy Number Analysis on the two cell lines revealed several regions of amplification and deletion, including a >10 copies amplification in chromosome 10 encompassing FGFR2 oncogene.

Gene	coding change	aminoacidic change	tumor FFPE	ccfDNA at diagnosis	ccfDNA at disease progression	cell line	PDX Passage1	predicted as pathogenic*
ALK	c.A1301G	p.K434R	not detected	not detected	7.0%	not detected	not detected	no
APC	4607insA	p.T1556fs*3	9.8%	22.6%	29.9%	100.0%	95.0%	ND
APC	c.A2659C	p.S887R	not detected	not detected	not detected	43.2%	not detected	yes
ARID1A	c.C3826T	p.R1276X	8.9%	28.5%	22.8%	53.0%	19.8%	yes
CREBBP	c.T1448C	p.L483P	not detected	not detected	not detected	not detected	21.7%	ND
DOT1L	c.A1193C	p.K398T	not detected	not detected	not detected	38.2%	not detected	yes
EP300	c.A7118C	p.N2373T	not detected	not detected	not detected	not detected	32.2%	no
EPHA5	c.A1417C	p.T473P	not detected	not detected	8.3%	not detected	not detected	no
ERBB3	c.C384A	p.S128R	9.8%	20.7%	20.4%	41.09%	48.60%	yes
ERBB4	c.T343A	p.Y115N	not detected	not detected	not detected	not detected	33.8%	yes
FAT1	c.A2084C	p.N695T	not detected	not detected	2.6%	not detected	not detected	no
GRIN2A	c.A4064C	p.K1355T	not detected	not detected	not detected	not detected	22.3%	no
KDM5A	c.T3547C	p.W1183R	not detected	not detected	not detected	not detected	22.2%	yes
KDR	c.A981C	p.K327N	not detected	not detected	not detected	not detected	27.9%	ND
KEAP1	c.G404A	p.R135H	9.0%	22.6%	19.5%	51.8%	44.0%	yes
KMT2C	c.C2459T	p.T820I	not detected	1.1%	1.1%	not detected	not detected	yes
KMT2C	c.C1013T	p.S338L	not detected	5.5%	4.3%	not detected	not detected	no
KMT2C	c.C2689T	p.R897X	not detected	not detected	1.4%	not detected	not detected	yes
KMT2C	c.C2228T	p.P743L	not detected	not detected	1.6%	not detected	not detected	no
MED12	c.A688C	p.I230L	not detected	not detected	not detected	not detected	25.0%	no
MGA	c.A395C	p.N132T	not detected	not detected	not detected	not detected	30.5%	yes
MGA	c.T4004C	p.L1335P	not detected	not detected	1.6%	not detected	not detected	ND
NF1	c.A6083G	p.K2028R	not detected	not detected	not detected	not detected	25.4%	yes
NTRK1	c.C2206T	p.Q736X	8.3%	26.0%	22.3%	32.6%	44.7%	yes
PALB2	C2419T	p.P807S	9.0%	19.1%	16.6%	not detected	47.4%	no
PTPRD	c.T2908G	p.L970V	not detected	5.1%	3.5%	not detected	47.7%	yes
PTPRD	c.A3117C	p.Q1039H	not detected	not detected	6.8%	not detected	not detected	yes
PTPRT	c.T479C	p.F160S	not detected	not detected	not detected	not detected	33.8%	no

RBM10	c.A2024G	p.K675R	not detected	not detected	not detected	not detected	41.7%	no
TP53	c.C9G	p.C3W	not detected	1.2%	not detected	not detected	not detected	yes
TP53	c.A355G	p.I119V	not detected	1.7%	not detected	not detected	not detected	yes
TP53	c.G315C	p.M105I	not detected	1.8%	not detected	not detected	not detected	yes
TP53	c.T316C	p.C106R	not detected	2.1%	not detected	not detected	not detected	yes
ZFHX3	c.A1604C	p.E535A	not detected	not detected	not detected	not detected	27.5%	no
ZFHX3	c.A7529G	p.Q2510R	not detected	not detected	not detected	not detected	39.7%	no

Table 7. Genetic alteration in CB055.

*variants were considered pathogenic when more than 50% of the 8 predictors (SIFT, Polyphen2 HVAR, LRT, Mutation Taster, Mutation Assessor, FATHMM, CADD and VEST) considered the alteration as pathogenic/damaging/deleterious/harmful. ND= not defined

Gene	**coding change**	**aminoacidic change**	**CUP#96 cell line**	**ccfDNA at diagnosis**	**predicted as pathogenic ***
PIK3C2G	c.G1813A	p.V605I	100.00%	48.35%	no
TP53	c.G418A	p.V140M	100.00%	49.34%	yes
SMAD4	c.A1610G	p.D537G	100.00%	44.23%	yes
KMT2C	c.G943A	p.G315S	4.58%	5.31%	yes
CTNNB1	c.C110T	p.S37F	55.10%	22.39%	yes
KMT2C	c.A2725G	p.R909G	-	6.48%	yes
KMT2C	c.C3274T	p.R1092X	-	5.28%	yes
KMT2C	c.C1013T	p.S338L	-	4.57%	no
KMT2C	c.G944A	p.G315D	-	3.09%	yes

Table 8. Genetic alteration in CB096.

*variants were considered pathogenic when more than 50% of the 8 predictors (SIFT, Polyphen2 HVAR, LRT, Mutation Taster, Mutation Assessor, FATHMM, CADD and VEST) considered the alteration as pathogenic/damaging/deleterious/harmful. ND= not defined

3.9 FGFR2 amplification and target therapy

FGFR2 amplification was found in both our patient (CB055 and CB096) and was confirmed in models' sample with Fluorescence In Situ Hybridization (FISH) and droplet digital PCR (ddPCR).

FISH demonstrated a different nature of FGFR2 amplification, in particular CB096 displays an extrachromosomal DNA amplification in the form of double minutes, while CB055 shows a homogeneously staining region (HSR) associated with FGFR2 chromosomal amplification (Fig. 5A). ddPCR confirmed the copy number variation (CNV) of FGFR2 gene using a probe-based assay in circulating cell-free (ccfDNA), cell line and PDX (Fig. 5B), with a copy number (CN) about 90 for CB055, and >400 for CB096.

This data provided the rationale for the pharmacological targeting of this receptor for therapeutic intervention.

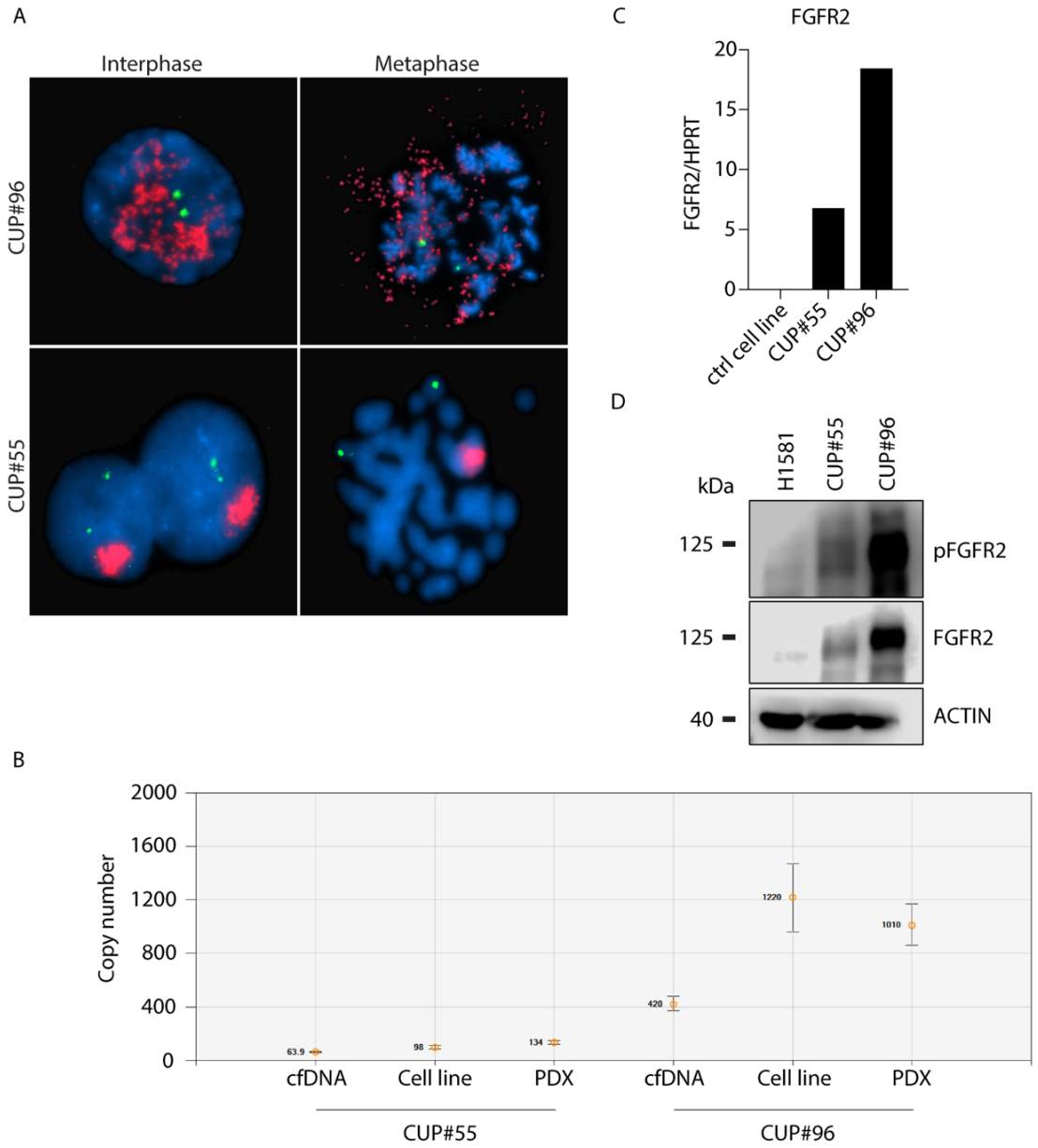

Figure 5. FGFR2 amplification in CUP models. A) FISH analysis shows the different nature of FGFR2 amplification in the two CUP cell models: double minutes for CB096 and HSR for CB055; FGFR2 gene probe is in red, Chr10 centromere probe in green, nucleus in blue. B) Copy number variation (CNV) analysis of FGFR2 gene detected by ddPCR using a probe-based assay. C) FGFR2 gene expression in the two amplified models and a control not amplified cell line. Bars represent the ratio between the gene copies and the reference HPRT copies assessed by ddPCR. D) Evaluation by western blot analysis of total and phosphorylated forms of FGFR2 in CB055 and CB096.

3.11 FGFR2 inhibitors therapy

We preliminarily tested a panel of multikinase inhibitors (BJ398, dovitinib, ponatinib) for their effect on CB055 cell vitality. We selected BGJ398 (infigratinib), a selective inhibitor of the FGFR family, for its greater antitumor activity.

In CB055, the treatment with this drug at different concentrations (range 0.5-2.5 μM) inhibited FGFR2 phosphorylation and the downstream AKT/mTOR/p70S6K signaling (Fig. 6A). In contrast, ERK1/2 remained phosphorylated even at the highest BGJ398 concentration. CB055 cells were sensitive to BGJ398 treatment, although a complete inhibition of their viability/proliferation was not achieved even at 1 μM BGJ398 concentration, suggesting that FGFR-independent mechanisms sustain the activation of the MAPK pathway and contribute to cell growth in this model.

These results prompted us to test whether targeting the MAPK pathway with trametinib, a highly specific inhibitor of MEK1/2 proteins, might improve the efficacy of BGJ398 treatment.

The drug combination inhibited both the AKT and MAPK pathways (Fig. 6B), had a highly significant synergistic inhibition of cell proliferation, as indicated by the comparison with the theoretical interaction curve in the Bliss experimental model ($p<0.001$) (Fig. 6C).

Interestingly, even if the single treatment with either BGJ398 or trametinib did not induce cell death in CB055 cells, their combination had a cytotoxic effect, as demonstrated by fluorescence microscopy analysis of Hoechst 33342/PI-stained cells; the morphology of the stained nuclei suggested that the cells died by apoptosis (Fig. 6D).

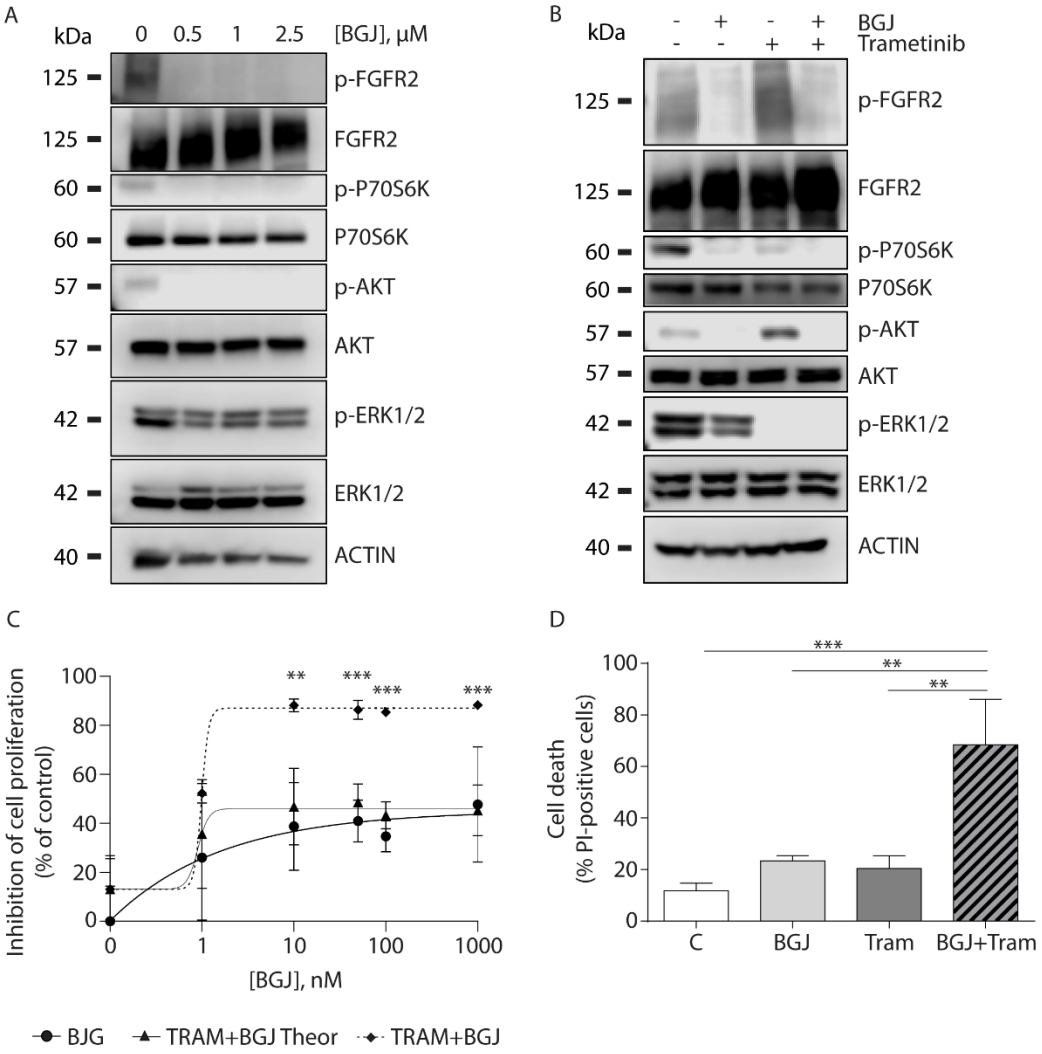

Figure 6. In vitro FGFR2 targeting and synergic activity of FGFR2 and MEK inhibitors in CB055. A) CB055 cells were treated with BGJ398 at the indicated concentrations; after 24h, protein extracts were analyzed by Western blotting for the indicated proteins. The results are representative of two independent experiments; B) CB055 cells were incubated with BGJ398 1µM, trametinib 100 nM, or the combination; after 24h, protein extracts were analyzed by Western blotting for the indicated proteins. The results are representative of two independent experiments; C) CB055 cells were treated with increasing concentrations of BGJ398 in combination with trametinib 100 nM. After 96h, cell proliferation was assessed by MTS assay. The data are expressed as percent inhibition vs. control. The asterisks indicate the statistical significance vs. the corresponding points of the Bliss Theoretical curve. The results are representative of three independent experiments; D) CB055 cells were incubated with BGJ398 1µM and/or trametinib 100nM; after 96h, the percentage of cell death was evaluated by fluorescence microscopy after Hoechst 33342/PI staining. The data are mean values ±SD of three independent experiments. **p<0.01, ***p<0.001.

Then, we analyzed the effects in CB096 cells. Inhibition of FGFR2 phosphorylation by BGJ398 treatment resulted in the downregulation of phosphorylated forms of AKT, p70S6K, and ERK1/2 levels, suggesting that both AKT and MAPK pathways are downstream of FGFR2 in this cell model (Fig. 7A). We tested the combination of BGJ398 with trametinib also in these cells and found that it produced a remarkable synergistic inhibition of cell proliferation (Fig. 7B-C), comparable to that observed in the other patient cells, suggesting that trametinib can provide a more effective inhibition of the MAPK pathway leading to suppression of cell growth also in this cell model. Of note, BGJ398 and trametinib alone were cytotoxic in CB096 cells, an effect that was further enhanced by the combined treatment (Fig. 7D).

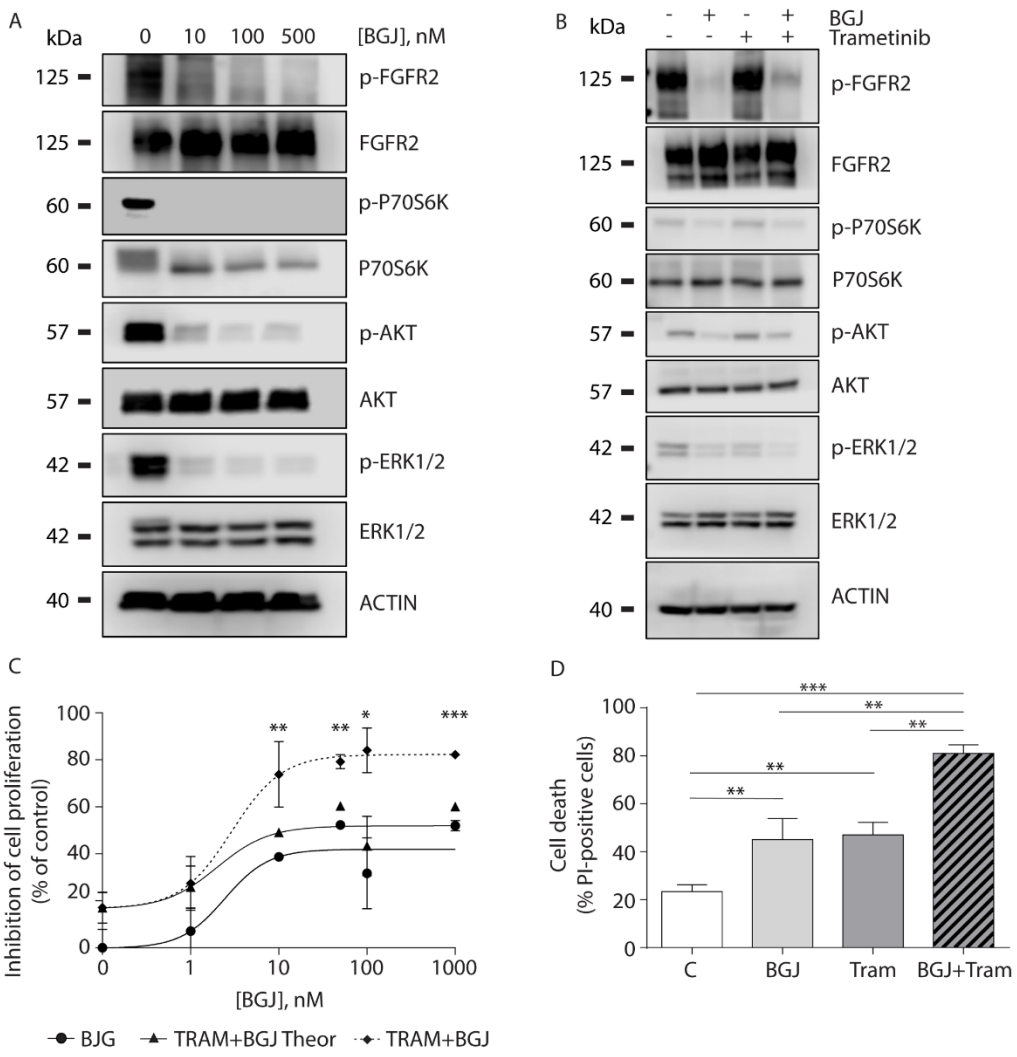

Figure 7. In vitro FGFR2 targeting and synergic activity of FGFR2 and MEK inhibitors in CB096. A) CB096 cells were treated with BGJ398 at the indicated concentrations; after 24h, protein extracts were analyzed by Western blotting for the indicated proteins. The results are representative of two independent experiments; B) CB096 cells were incubated with BGJ398 100nM, trametinib 10 nM or the combination; after 24h, protein extracts were analyzed by Western blotting for the indicated proteins. The results are representative of two independent experiments; C) CB096 cells were treated with increasing concentrations of BGJ398 in combination with trametinib 10 nM. After 96h, cell proliferation was assessed by MTS assay. The data are expressed as percent inhibition vs. control. The asterisks indicate the statistical significance vs. the corresponding points of the Bliss Theoretical curve. The results are representative of three independent experiments; D) CB096 cells were incubated with BGJ398 100 nM and/or trametinib 10 nM; after 96h, the percentage of cell death was evaluated by fluorescence microscopy after Hoechst 33342/PI staining. The data are mean values ±SD of three independent experiments. *p<0.05; **p<0.01, ***p<0.001, ****p < 0.0001.

3.12 Synergic activity of FGFR2 and MEK inhibitors in vivo

We subsequently tested whether in vitro findings on FGFR2 and MEK inhibitors synergic activity in CUP#55 and CUP#96 PDX models in vivo.

Mice were treated with control vehicle, BGJ398 (15 mg/kg), trametinib (0.6 mg/kg), and trametinib/BGJ398 combination (N=5 per group) as reported in Fig. 8A and B, and the effect of different drugs was evaluated measuring the tumor volume.

Single drug (BGJ398 or trametinib) treatment significantly reduced the tumor volume in PDX#CB096 but not in PDX#CB055 if compared to tumors of mice treated with vehicle.

However, the combination of trametinib and BGJ398 resulted more effective than the single treatments in both PDX models, as demonstrated by the dramatic reduction of the tumor volume (Fig. 8A, B).

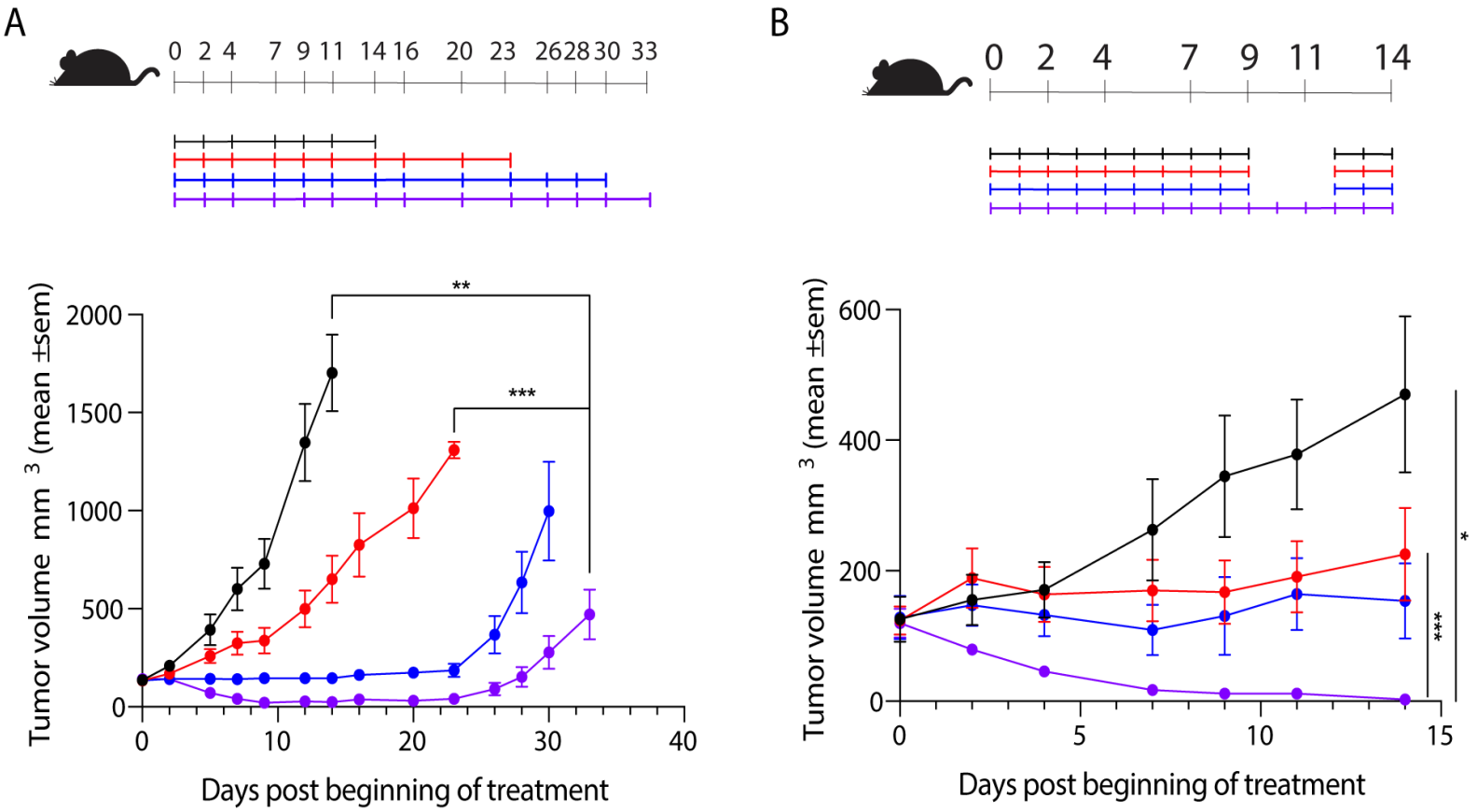

Figure 8. In vivo activity of FGFR2 and MEK inhibitors. In vivo treatment of A) PDX#CB096 PDXs and B) PDX#CB055 PDXs. The timeline shows the day of tumor volume measurement, and the color bars represent the scheme of each treatment. The line graph shows that the combination of trametinib+BGJ398 treatments (purple) is more effective than the single ones (trametinib in red and BGJ398 in blue) in both PDX models. N=5 mice per group. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

3.13 Cell type identification using single cell transcriptomics

Single-cell gene expression profiles deriving from CB055 cell line were clustered on the basis of their reciprocal similarity, leading to the identification of two separate clusters (Figure 9).

We then applied a mapping approach using an atlas of liver cells as a reference, with the objective of inferring the most likely type of origin for CUP cells. While 93% of cells in cluster 1 were annotated as cholangiocytes, the classification of cluster 2 was more diverse: specifically, mapped types were split between cholangiocytes (73% of cells) and T/NK cells (27%). As this separation did not correspond to any difference in cell morphology, we double-checked the association manually. We identified the marker genes for the T/NK subgroup, which included DIAPH3, CENPP, BRIP1, ASPM, FAF1, STAG1, POLA1, POLQ, MKI67 and CIT. This list is notable for two reasons: most of the genes are implicated in the cell cycle, and in particular with the mitotic phase. Furthermore, the list lacks any of the markers which characterize the T/NK group in the liver atlas, such as CCL5, KLRB1, NKG7, CD69, GZMA, CCL4, CST7, GNLY, GZMK. None of the cells in cluster 2 expresses any of those genes. This suggests that the identification of T/NK cells by reference mapping is likely incorrect, and we hypothesize this result is driven by the confounding effect of the cell cycle. An average of 46% of the cells in the reference atlas are actively cycling, but when we focus on T/NK cells alone the percentage rises to 74%. This imbalance makes it likely that cycling cells from our own sample are erroneously recognized as belonging to the T/NK population of the reference.

3.14 Expression profiles of cell clusters

The transcriptome analysis highlighted the presence of two separate cell clusters (Figure 9). composed of transformed cholangiocytes (CB055 sub-population 1) with significant differences in their expression profiles. We identified two sets of positive marker genes which characterize each cluster. Specifically, we obtained a total of 1051 markers for the first cluster and 3059 for the second.

Functional annotation of the markers for cluster 1 highlighted an over-representation for pathways including “Eukaryotic Translation Elongation” (Reactome, FDR 1E-89), “Electron transport chain: OXPHOS system in mitochondria” (WikiPathways, FDR 3E-60), “Vif-mediated degradation of APOBEC3G” (Reactome, FDR 9.3E-12), and “PTEN Regulation” (Reactome, FDR 5E-6).

A similar analysis of cluster 2 markers identified the following pathways: “Signaling by Rho GTPases, Miro GTPases and RHOBTB3” (Reactome, FDR 4E-27), “Chromatin modifying enzymes” (Reactome, FDR 4.9E-10), “NRAGE signals death through JNK” (Reactome, FDR 1.5E-4), “Signaling

by TGF-beta Receptor Complex" (WikiPathways, FDR 3.7E-4) and "Signaling by cytosolic FGFR1 fusion mutants" (Reactome, FDR 4E-3).

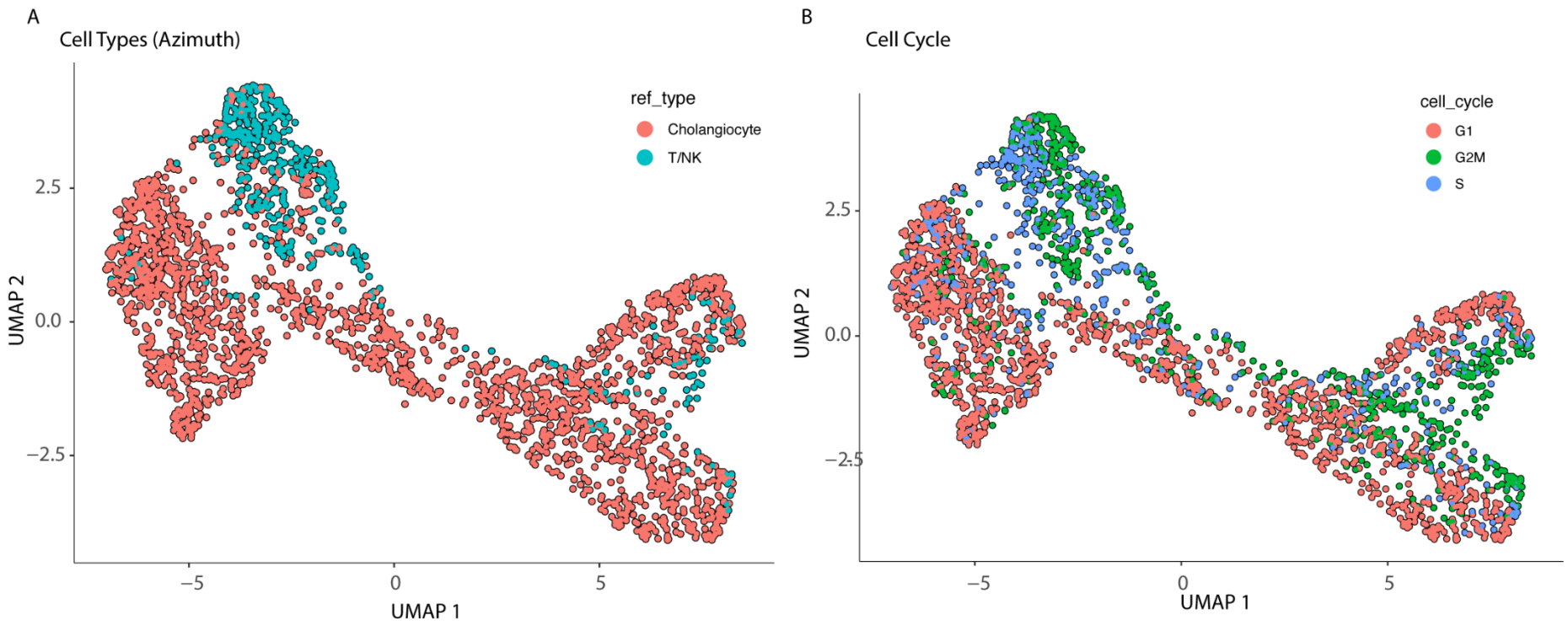

Figure 9. Separation of Cells clusters.

4 DISCUSSION

4.1 MicroRNA profiling

The identification of origin in metastatic cancers is well known for its complexity. The diagnostic algorithm requires careful collaborative work between the physician and the pathologist, 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 identifiable, despite an intensive clinical and pathological diagnostic workup (Fizazi et al. 2015). CUPs represent an enigma for its biological and pathological behaviour, and an important under-researched clinical problem.

In the past years, many molecular tests have been proposed. In particular based on gene expression (GEP), microRNA or DNA methylation profile, for the purpose to improve primary site identification.

Molecular profiling assays are based on the preservation of molecular signature of the primary site in metastatic tumour. Overall the molecular predictors can reach an accuracy from 80 to 95% (Conway et al. 2019). These promising results 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.

We developed a molecular assay to assess the expression of 89 miRNAs in tumour FFPE samples by using Droplet Digital PCR (ddPCR).

Our panel was developed merging two cancer-specific miRNA signatures previously identified in two microarray-based studies (Ferracin et al. 2011; Rosenfeld et al. 2008). 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 (Hindson et al. 2013). In addition, EvaGreen-based ddPCR allows to precisely detect target miRNAs at levels down to 1 copy/µL (Miotto et al. 2014).

As hypothesized, this approach is very convenient since we were able to successfully analyse the totality of FFPE samples in our cohort (100% success rate), with no excluded sample due to technical issues.

In this part of the study we analysed the 89-miRNA profiles of 156 FFPE samples, including 50 CUPs, and successfully obtained a primary site prediction for all patients. We also observed an interesting consistency between our prediction outcomes and prior clinical and histopathological hypotheses.

Frequency rates of predicted tissue of origin is coherent with those reported in literature. The most common primary sites where found post-mortem in lung (27%), pancreas (24%), liver or bile duct system (8%), kidney or adrenal (8%) or colon (7%) (Pentheroudakis, Golfinopoulos, and Pavlidis 2007). Of note, some CUPs were molecularly predicted as LUAD with a negativity for TTF1, which defines a subgroup of LUAD with unfavorable outcomes (Zhang et al. 2015).

Three molecular assays were recently approved for CUP diagnostics in US: Pathwork Tissue of Origin Test (Pathwork Diagnostics, Redwood City, CA, USA), CancerTYPE ID (bioTheranostics, San Diego, CA, USA), and miRview mets2 (Rosetta Genomics, North Brunswick, NJ, USA). The first is a microarray-based system to assess the gene expression profiles (GEP) of 2000 genes claim to distinguish up to 15 tumour 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 tumour types. Finally, miRview mets^2 system, assessing the expression of 64 miRNAs by RT-qPCR, is able to distinguish up to 26 tumour types.

Our ddPCR tool has the advantage to cover a high number of tumour classes (with set 16 primary tumour types and 19 subtypes), selected as the most common CUP tissues of origin. Our assay has a 100% success rate and requires a 2-day 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.

4.2 CUP model's treatment

Treatment choice for cancer of unknown primary patients is always challenging due to the inconclusiveness of classical diagnostic investigations in tumour type identification. In this scenario, therapy can be based on empirical approaches, molecular predictions of the primary site, the clinical assessment of the similarity with other known tumour types or more recently, on personalized genetic approaches. Indeed, the combination of molecular and genetic investigations can help the treating clinicians to define the most probable site-of-origin or potential druggable mutations and use this information to choose the best therapeutic strategies.

CUP therapeutic choice has been also hampered by the lack of CUP models on which to test novel therapeutic approaches or investigate the biology of this aggressive disease.

The first CUP experimental model was described by Verginelli et al. (Verginelli et al. 2021). The authors demonstrated that CUP models present a specific stem-cell like phenotype and tested for the first time their sensitivity to MEK inhibitors.

In this part of the study, we described two in vitro and in vivo models we derived from ascites circulating tumours cells (CTCs) of CUP patients, named CB055 and CB096. Using a CTC-optimized protocol, we established two long-term cell cultures spontaneously growing as spheroids/tumoroids and expressing stem-cell markers. The ascites CTCs were engrafted into mice to generate two patients derived xenografts, which recapitulated the characteristics of the original tumour.

The genetic analyses, performed on tumour tissues and models, revealed that the two models shared the FGFR2 amplification as main genetic alteration, although of a different nature. FGFR2 amplification is reported in several solid tumours, including gastric cancer and breast cancer (Katoh 2008), while its translocation is a recurrent feature in cholangiocarcinoma (Smyth, Babina, and Turner 2017).

FGFR2 amplification is associated with increased levels of the protein and its aberrant phosphorylation leads to the activation of downstream pathways, including MAPK-ERK signalling (Babina and Turner 2017) which in turn accelerate cell proliferation. Since FGFR2 amplification is a druggable target (Babina and Turner 2017), the treatment with BGJ398 (infigratinib), a pan-FGFR inhibitor, demonstrated the effectiveness of this target therapy in reducing AKT activation, but did not completely prevent ERK1/2 phosphorylation, which remained active through alternative mechanisms. This finding suggests that FGFR2 targeting in FGFR2-amplified CUPs might not be sufficient to halt tumour growth, due to the concomitant activation of alternative survival/proliferation pathways. Therefore, we investigated the effect of the combined use of trametinib, a MAPK pathway inhibitor selective for MEK1/2. MAPK is a pathway that is frequently activated in CUPs (Krikelis et al. 2012), where its activation correlates with worse prognosis. Interestingly, the combined treatment with FGFR2 and MEK inhibitors generated a remarkable synergistic effect, reducing cell growth and viability in both cell models. To validate the results in a preclinical model, the corresponding PDXs were treated with the drug combination and the effect was a drastic reduction of the tumour size if compared with single treatments. The treatment

combination was well tolerated by the CB096 model but aggravated the mice frailty in CB055 model, which already presented signed of tumor-induced cachexia.

In the last years, pemigatinib and infigratinib, both orally active agents targeting FGFR1-4, have received an accelerated approval by US Food and Drug Administration (FDA) for the treatment of adult patients with previously treated and unresectable or metastatic cholangiocarcinoma harboring FGFR2 fusion or other rearrangement. Based on some concerns, European agency (EMA) granted approval with the same indication only for pemigatinib. Likewise, erdafitinib, an orally active small potent TKI of FGFR1–4, was granted accelerated approval for patients with locally advanced or metastatic urothelial carcinoma, with susceptible FGFR3 or FGFR2 genetic alterations.

FGFR inhibitors have also been largely tested in gastric cancer patients with less enthusiastic results so far. However, positive results from a phase II trial have been reported by using bemarituzumab in addition to chemotherapy as first line treatment for metastatic HER-2 negative gastroesophageal cancer patients with FGFR2b hyperexpression or FGFR2 gene amplification (Wainberg et al. 2022). As such, several phase II and phase III trials currently ongoing. Other new compounds, alone or in combination with other agents, are under investigation for the treatment of multiple solid tumours carrying FGFR alterations. Our results suggest that the combination of FGFR and MEK inhibitors could be a potential strategy to improve clinical outcomes in cancer patients carrying FGFR gene alterations. This means that co-targeting cross-talking pathways may potentiate FGFR inhibition, and improve the therapeutic benefit, as we have demonstrated with the MEK inhibitor trametinib.

Based on the evidence that MEK inhibition in KRAS-mutant lung cancer leads to compensatory MAPK pathway reactivation through FGFR1, combining trametinib with FGFR1-specific inhibitors encapsulated in nanoparticles allowed to efficaciously inhibit growth and proliferation in KRAS-mutant/FGFR compensatory cancer cells.

5 CONCLUSIONS

Our results provide further evidences 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 tumours. 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 second part of the study contributes to fill the gap in CUP model availability, through the generation of two human cell lines and corresponding PDXs. We demonstrated how the molecular characterization and genetic profile of the tumour can provide information to predict the most probable site-of-origin and identify druggable targets. In fact, the development of models that of human tumours in vitro and in vivo, has become a helpful tool for drug screening, particularly to assess new therapeutic combinations that could be translated to patients with the same genetic alterations or molecular features.

In conclusion, our study demonstrated that digital miRNA expression profiling of CUP samples has the potential to be employed in a clinical setting. 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. Finally, the availability of CUP models for future studies will contribute to deepen our knowledge on the mechanisms at the base of CUP high proliferative and metastatic potential and still mysterious biology.

6 BIBLIOGRAPHY

AC, D. E. Miguel, R. P. Cid, J. M. Trufero, I. P. Bernad, L. C. Urquizu, J. H. Cubero, and A. A. Torres. 2015. 'Gene expression profiling and its use in adenocarcinomas of unknown primary origin: A case report', *Oncol Lett*, 10: 2657-61.

Al-Brahim, N., C. Ross, B. Carter, and K. Chorneyko. 2005. 'The value of postmortem examination in cases of metastasis of unknown origin-20-year retrospective data from a tertiary care center', *Ann Diagn Pathol*, 9: 77-80.

Babina, I. S., and N. C. Turner. 2017. 'Advances and challenges in targeting FGFR signalling in cancer', *Nat Rev Cancer*, 17: 318-32.

Blaszyk, H., A. Hartmann, and J. Bjornsson. 2003. 'Cancer of unknown primary: clinicopathologic correlations', *APMIS*, 111: 1089-94.

Briasoulis, E., G. Fountzilas, A. Bamias, M. A. Dimopoulos, N. Xiros, G. Aravantinos, E. Samantas, H. Kalofonos, T. Makatsoris, N. Mylonakis, P. Papakostas, D. Skarlos, I. Varthalitis, and N. Pavlidis. 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: 277-84.

Bridgewater, J., R. van Laar, A. Floore, and T. Veer L. Van. 2008. 'Gene expression profiling may improve diagnosis in patients with carcinoma of unknown primary', *Br J Cancer*, 98: 1425-30.

Chaffer, C. L., and R. A. Weinberg. 2011. 'A perspective on cancer cell metastasis', *Science*, 331: 1559-64.

Chalmers, Z. R., C. F. Connelly, D. Fabrizio, L. Gay, S. M. Ali, R. Ennis, A. Schrock, B. Campbell, A. Shlien, J. Chmielecki, F. Huang, Y. He, J. Sun, U. Tabori, M. Kennedy, D. S. Lieber, S. Roels, J. White, G. A. Otto, J. S. Ross, L. Garraway, V. A. Miller, P. J. Stephens, and G. M. Frampton. 2017. 'Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden', *Genome Med*, 9: 34.

Chu, P., E. Wu, and L. M. Weiss. 2000. 'Cytokeratin 7 and cytokeratin 20 expression in epithelial neoplasms: a survey of 435 cases', *Mod Pathol*, 13: 962-72.

Clynick, B., B. Dessauvagie, G. Sterrett, N. T. Harvey, R. J. N. Allcock, C. Saunders, W. Erber, and K. Meehan. 2018. 'Genetic characterisation of molecular targets in carcinoma of unknown primary', *J Transl Med*, 16: 185.

Conway, A. M., C. Mitchell, E. Kilgour, G. Brady, C. Dive, and N. Cook. 2019. 'Molecular characterisation and liquid biomarkers in Carcinoma of Unknown Primary (CUP): taking the 'U' out of 'CUP'', *Br J Cancer*, 120: 141-53.

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

Didolkar, M. S., N. Fanous, E. G. Elias, and R. H. Moore. 1977. 'Metastatic carcinomas from occult primary tumors. A study of 254 patients', *Ann Surg*, 186: 625-30.

Dumur, C. I., M. Lyons-Weiler, C. Sciulli, C. T. Garrett, I. Schrijver, T. K. Holley, J. Rodriguez-Paris, J. R. Pollack, J. L. Zehnder, M. Price, J. M. Hagenkord, C. T. Rigl, L. J. Buturovic, G. G. Anderson, and F. A. Monzon. 2008. 'Interlaboratory performance of a microarray-based gene expression test to determine tissue of origin in poorly differentiated and undifferentiated cancers', *J Mol Diagn*, 10: 67-77.

Duval, J. V., L. Savas, and B. F. Banner. 2000. 'Expression of cytokeratins 7 and 20 in carcinomas of the extrahepatic biliary tract, pancreas, and gallbladder', *Arch Pathol Lab Med*, 124: 1196-200.

Fernandez, A. F., Y. Assenov, J. I. Martin-Subero, B. Balint, R. Siebert, H. Taniguchi, H. Yamamoto, M. Hidalgo, A. C. Tan, O. Galm, I. Ferrer, M. Sanchez-Cespedes, A. Villanueva, J. Carmona, J. V. Sanchez-Mut, M. Berdasco, V. Moreno, G. Capella, D. Monk, E. Ballestar, S. Ropero, R. Martinez, M. Sanchez-Carbayo, F. Prosper, X. Agirre, M. F. Fraga, O. Grana, L. Perez-Jurado, J. Mora, S. Puig, J. Prat, L. Badimon, A. A. Puca, S. J. Meltzer, T. Lengauer, J. Bridgewater, C. Bock, and M. Esteller. 2012. 'A DNA methylation fingerprint of 1628 human samples', *Genome Res*, 22: 407-19.

Ferracin, M., and M. Negrini. 2015. 'Micromarkers 2.0: an update on the role of microRNAs in cancer diagnosis and prognosis', *Expert Rev Mol Diagn*, 15: 1369-81.

Ferracin, M., M. Pedriali, A. Veronese, B. Zagatti, R. Gafa, E. Magri, M. Lunardi, G. Munerato, G. Querzoli, I. Maestri, L. Ulazzi, I. Nenci, C. M. Croce, G. Lanza, P. Querzoli, and M. Negrini. 2011. 'MicroRNA profiling for the identification of cancers with unknown primary tissue-of-origin', *J Pathol*, 225: 43-53.

Ferracin, M., A. Veronese, and M. Negrini. 2010. 'Micromarkers: miRNAs in cancer diagnosis and prognosis', *Expert Rev Mol Diagn*, 10: 297-308.

Fizazi, K., F. A. Greco, N. Pavlidis, G. Daugaard, K. Oien, G. Pentheroudakis, and Esmo Guidelines Committee. 2015. 'Cancers of unknown primary site: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up', *Ann Oncol*, 26 Suppl 5: v133-8.

Fumarola, C., S. La Monica, R. R. Alfieri, E. Borra, and G. G. Guidotti. 2005. 'Cell size reduction induced by inhibition of the mTOR/S6K-signaling pathway protects Jurkat cells from apoptosis', *Cell Death Differ*, 12: 1344-57.

Gatalica, Z., S. Z. Millis, S. Vranic, R. Bender, G. D. Basu, A. Voss, and D. D. Von Hoff. 2014. 'Comprehensive tumor profiling identifies numerous biomarkers of drug response in cancers of unknown primary site: analysis of 1806 cases', *Oncotarget*, 5: 12440-7.

Gatalica, Z., J. Xiu, J. Swensen, and S. Vranic. 2018. 'Comprehensive analysis of cancers of unknown primary for the biomarkers of response to immune checkpoint blockade therapy', *Eur J Cancer*, 94: 179-86.

Gevaert, O., A. Daemen, B. De Moor, and L. Libbrecht. 2009. 'A taxonomy of epithelial human cancer and their metastases', *BMC Med Genomics*, 2: 69.

Golfinopoulos, V., G. Pentheroudakis, S. Kamakari, V. Metaxa-Mariatou, and N. Pavlidis. 2009. 'Donor-derived breast cancer in a bone marrow transplantation recipient', *Breast Cancer Res Treat*, 113: 211-3.

Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. 1999. 'Molecular classification of cancer: class discovery and class prediction by gene expression monitoring', *Science*, 286: 531-7.

Greco, F. A., W. J. Lennington, D. R. Spigel, and J. D. Hainsworth. 2013. 'Molecular profiling diagnosis in unknown primary cancer: accuracy and ability to complement standard pathology', *J Natl Cancer Inst*, 105: 782-90.

Greco, F. A. 2015. 'Poorly differentiated neoplasms of unknown primary site: diagnostic usefulness of a molecular cancer classifier assay', *Mol Diagn Ther*, 19: 91-7.

Greco, F. A., and N. Pavlidis. 2009. 'Treatment for patients with unknown primary carcinoma and unfavorable prognostic factors', *Semin Oncol*, 36: 65-74.

Grewal, J. K., B. Tessier-Cloutier, M. Jones, S. Gakkhar, Y. Ma, R. Moore, A. J. Mungall, Y. Zhao, M. D. Taylor, K. Gelmon, H. Lim, D. Renouf, J. Laskin, M. Marra, S. Yip, and S. J. M. Jones. 2019. 'Application of a Neural Network Whole Transcriptome-Based Pan-Cancer Method for Diagnosis of Primary and Metastatic Cancers', *JAMA Netw Open*, 2: e192597.

Groschel, S., M. Bommer, B. Hutter, J. Budczies, D. Bonekamp, C. Heining, P. Horak, M. Frohlich, S. Uhrig, D. Hubschmann, C. Georg, D. Richter, N. Pfarr, K. Pfutze, S. Wolf, P. Schirmacher, D. Jager, C. von Kalle, B. Brors, H. Glimm, W. Weichert, A. Stenzinger, and S. Frohling. 2016. 'Integration of genomics and histology revises diagnosis and enables effective therapy of refractory cancer of unknown primary with PDL1 amplification', *Cold Spring Harb Mol Case Stud*, 2: a001180.

Gross-Goupil, M., A. Fourcade, E. Blot, N. Penel, S. Negrier, S. Culine, L. Chaigneau, T. Lesimple, F. Priou, A. Lortholary, M. C. Kaminsky, J. Provencal, E. Voog, J. Bouzy, A. Laplanche, and K. Fizazi. 2012. 'Cisplatin alone or combined with gemcitabine in carcinomas of unknown primary: results of the randomised GEFCAPI 02 trial', *Eur J Cancer*, 48: 721-7.

Hainsworth, J. D., G. Daugaard, T. Lesimple, G. Hubner, F. A. Greco, M. J. Stahl, C. M. Buschenfelde, D. Allouache, N. Penel, P. Knoblauch, and K. S. Fizazi. 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: 1654-61.

Hainsworth, J. D., and K. Fizazi. 2009. 'Treatment for patients with unknown primary cancer and favorable prognostic factors', *Semin Oncol*, 36: 44-51.

Hainsworth, J. D., M. S. Rubin, D. R. Spigel, R. V. Boccia, S. Raby, R. Quinn, and F. A. Greco. 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: 217-23.

Hainsworth, J. D., C. A. Schnabel, M. G. Erlander, D. W. Haines, 3rd, and F. A. Greco. 2012. 'A retrospective study of treatment outcomes in patients with carcinoma of unknown primary site and a colorectal cancer molecular profile', *Clin Colorectal Cancer*, 11: 112-8.

Hamilton, C. S., and A. O. Langlands. 1987. 'ACUPS (adenocarcinoma of unknown primary site): a clinical and cost benefit analysis', *Int J Radiat Oncol Biol Phys*, 13: 1497-503.

Handorf, C. R., A. Kulkarni, J. P. Grenert, L. M. Weiss, W. M. Rogers, O. S. Kim, F. A. Monzon, M. Halks-Miller, G. G. Anderson, M. G. Walker, R. Pillai, and W. D. Henner. 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: 1067-75.

Hayashi, H., T. Kurata, Y. Takiguchi, M. Arai, K. Takeda, K. Akiyoshi, K. Matsumoto, T. Onoe, H. Mukai, N. Matsubara, H. Minami, M. Toyoda, Y. Onozawa, A. Ono, Y. Fujita, K. Sakai, Y. Koh, A. Takeuchi, Y. Ohashi, K. Nishio, and K. Nakagawa. 2019. '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*, 37: 570-79.

Hindson, C. M., J. R. Chevillet, H. A. Briggs, E. N. Gallichotte, I. K. Ruf, B. J. Hindson, R. L. Vessella, and M. Tewari. 2013. 'Absolute quantification by droplet digital PCR versus analog real-time PCR', *Nat Methods*, 10: 1003-5.

Horlings, H. M., R. K. van Laar, J. M. Kerst, H. H. Helgason, J. Wesseling, J. J. van der Hoeven, M. O. Warmoes, A. Floore, A. Witteveen, J. Lahti-Domenici, A. M. Glas, L. J. Van't Veer, and D. de Jong. 2008. 'Gene expression profiling to identify the histogenetic origin of metastatic adenocarcinomas of unknown primary', *J Clin Oncol*, 26: 4435-41.

Huebner, G., H. Link, C. H. Kohne, M. Stahl, A. Kretzschmar, S. Steinbach, G. Folprecht, H. Bernhard, S. E. Al-Batran, P. Schoffski, C. Burkart, F. Kullmann, B. Otremba, M. Menges, M. Hoffmann, U. Kaiser, A. Aldaoud, A. Jahn, and C. U. P. Study Group German. 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: 44-9.

Jordan, W. E., 3rd, and R. A. Shildt. 1985. 'Adenocarcinoma of unknown primary site. The Brooke Army Medical Center experience', *Cancer*, 55: 857-60.

Kaaks, R., D. Sookthai, K. Hemminki, A. Kramer, H. Boeing, E. Wirfalt, E. Weiderpass, K. Overvad, A. Tjonneland, A. Olsen, P. H. Peeters, H. B. Bueno-de-Mesquita, S. Panico, V. Pala, P. Vineis, J. R. Quiros, E. Ardanaz, M. J. Sanchez, M. D. Chirlaque, N. Larranaga, P. Brennan, D. Trichopoulos, A. Trichopoulou, P. Lagiou, G. Hallmans, K. T. Khaw, T. J. Key, E. Riboli, and F. Canzian. 2014. 'Risk factors for cancers of unknown primary site: Results from the prospective EPIC cohort', *Int J Cancer*, 135: 2475-81.

Kato, S., N. Krishnamurthy, K. C. Banks, P. De, K. Williams, C. Williams, B. Leyland-Jones, S. M. Lippman, R. B. Lanman, and R. Kurzrock. 2017. 'Utility of Genomic Analysis In Circulating Tumor DNA from Patients with Carcinoma of Unknown Primary', *Cancer Res*, 77: 4238-46.

Katoh, M. 2008. 'Cancer genomics and genetics of FGFR2 (Review)', *Int J Oncol*, 33: 233-7.

Kerr, S. E., C. A. Schnabel, P. S. Sullivan, Y. Zhang, V. Singh, B. Carey, M. G. Erlander, W. E. Highsmith, S. M. Dry, and E. F. Brachtel. 2012. 'Multisite validation study to determine performance characteristics of a 92-gene molecular cancer classifier', *Clin Cancer Res*, 18: 3952-60.

Kodack, D. P., A. F. Farago, A. Dastur, M. A. Held, L. Dardaei, L. Friboulet, F. von Flotow, L. J. Damon, D. Lee, M. Parks, R. Dicecca, M. Greenberg, K. E. Kattermann, A. K. Riley, F. J. Fintelmann, C. Rizzo, Z. Piotrowska, A. T. Shaw, J. F. Gainor, L. V. Sequist, M. J. Niederst, J. A. Engelman, and C. H. Benes. 2017. 'Primary Patient-Derived Cancer Cells and Their Potential for Personalized Cancer Patient Care', *Cell Rep*, 21: 3298-309.

Krikelis, D., G. Pentheroudakis, A. Goussia, V. Siozopoulou, M. Bobos, D. Petrakis, A. Stoyianni, V. Golfinopoulos, A. Cervantes, T. Ciuleanu, G. Fountzilas, V. Malamou-Mitsi, and N. Pavlidis. 2012. 'Profiling immunohistochemical expression of NOTCH1-3, JAGGED1, cMET, and phospho-MAPK in 100 carcinomas of unknown primary', *Clin Exp Metastasis*, 29: 603-14.

La Monica, S., M. Galetti, R. R. Alfieri, A. Cavazzoni, A. Ardizzoni, M. Tiseo, M. Capelletti, M. Goldoni, S. Tagliaferri, A. Mutti, C. Fumarola, M. Bonelli, D. Generali, and P. G. Petronini. 2009. 'Everolimus restores gefitinib sensitivity in resistant non-small cell lung cancer cell lines', *Biochem Pharmacol*, 78: 460-8.

Laprovitera, N., M. Grzes, E. Porcellini, and M. Ferracin. 2018. 'Cancer Site-Specific Multiple microRNA Quantification by Droplet Digital PCR', *Front Oncol*, 8: 447.

Laprovitera, N., M. Riefolo, E. Porcellini, G. Durante, I. Garajova, F. Vasuri, A. Aigelsreiter, N. Dandachi, G. Benvenuto, F. Agostinis, S. Sabbioni, I. Berindan Neagoe, C. Romualdi, A.

Ardizzoni, D. Trere, M. Pichler, A. D'Errico, and M. Ferracin. 2021. 'MicroRNA expression profiling with a droplet digital PCR assay enables molecular diagnosis and prognosis of cancers of unknown primary', *Mol Oncol*, 15: 2732-51.

Laprovitera, N., I. Salamon, F. Gelsomino, E. Porcellini, M. Riefolo, M. Garonzi, P. Tononi, S. Valente, S. Sabbioni, F. Fontana, N. Manaresi, A. D'Errico, M. A. Pantaleo, A. Ardizzoni, and M. Ferracin. 2021. 'Genetic Characterization of Cancer of Unknown Primary Using Liquid Biopsy Approaches', *Front Cell Dev Biol*, 9: 666156.

Le Chevalier, T., E. Cvitkovic, P. Caille, J. Harvey, G. Contesso, M. Spielmann, and J. Rouesse. 1988. 'Early metastatic cancer of unknown primary origin at presentation. A clinical study of 302 consecutive autopsied patients', *Arch Intern Med*, 148: 2035-9.

Liu, A., M. T. Tetzlaff, P. Vanbelle, D. Elder, M. Feldman, J. W. Tobias, A. R. Sepulveda, and X. Xu. 2009. 'MicroRNA expression profiling outperforms mRNA expression profiling in formalin-fixed paraffin-embedded tissues', *Int J Clin Exp Pathol*, 2: 519-27.

Loffler, H., N. Pfarr, M. Kriegsmann, V. Endris, T. Hielscher, P. Lohneis, G. Folprecht, A. Stenzinger, M. Dietel, W. Weichert, and A. Kramer. 2016. 'Molecular driver alterations and their clinical relevance in cancer of unknown primary site', *Oncotarget*, 7: 44322-29.

Losa, F., G. Soler, A. Casado, A. Estival, I. Fernandez, S. Gimenez, F. Longo, R. Pazo-Cid, J. Salgado, and M. A. Segui. 2018. 'SEOM clinical guideline on unknown primary cancer (2017)', *Clin Transl Oncol*, 20: 89-96.

Ma, X. J., R. Patel, X. Wang, R. Salunga, J. Murage, R. Desai, J. T. Tuggle, W. Wang, S. Chu, K. Stecker, R. Raja, H. Robin, M. Moore, D. Baunoch, D. Sgroi, and M. Erlander. 2006. 'Molecular classification of human cancers using a 92-gene real-time quantitative polymerase chain reaction assay', *Arch Pathol Lab Med*, 130: 465-73.

Maiche, A. G. 1993. 'Cancer of unknown primary. A retrospective study based on 109 patients', *Am J Clin Oncol*, 16: 26-9.

Massard, C., Y. Loriot, and K. Fizazi. 2011. 'Carcinomas of an unknown primary origin--diagnosis and treatment', *Nat Rev Clin Oncol*, 8: 701-10.

Masuda, N., T. Ohnishi, S. Kawamoto, M. Monden, and K. Okubo. 1999. 'Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples', *Nucleic Acids Res*, 27: 4436-43.

Mayordomo, J. I., J. M. Guerra, C. Guijarro, M. D. Garcia-Prats, A. Gomez, M. Lopez-Brea, R. Gonzalez, P. Hergueta, M. A. Lopez-Pino, F. Martinez-Tello, and et al. 1993. 'Neoplasms of unknown primary site: a clinicopathological study of autopsied patients', *Tumori*, 79: 321-4.

Meiri, E., W. C. Mueller, S. Rosenwald, M. Zepeniuk, E. Klinke, T. B. Edmonston, M. Werner, U. Lass, I. Barshack, M. Feinmesser, M. Huszar, F. Fogt, K. Ashkenazi, M. Sanden, E. Goren, N. Dromi, O. Zion, I. Burnstein, A. Chajut, Y. Spector, and R. Aharonov. 2012. 'A second-generation microRNA-based assay for diagnosing tumor tissue origin', *Oncologist*, 17: 801-12.

Miotto, E., E. Saccenti, L. Lupini, E. Callegari, M. Negrini, and M. Ferracin. 2014. 'Quantification of circulating miRNAs by droplet digital PCR: comparison of EvaGreen- and TaqMan-based chemistries', *Cancer Epidemiol Biomarkers Prev*, 23: 2638-42.

Monzon, F. A., M. Lyons-Weiler, L. J. Buturovic, C. T. Rigl, W. D. Henner, C. Sciulli, C. I. Dumur, F. Medeiros, and G. G. Anderson. 2009. 'Multicenter validation of a 1,550-gene expression profile for identification of tumor tissue of origin', *J Clin Oncol*, 27: 2503-8.

Monzon, F. A., F. Medeiros, M. Lyons-Weiler, and W. D. Henner. 2010. 'Identification of tissue of origin in carcinoma of unknown primary with a microarray-based gene expression test', *Diagn Pathol*, 5: 3.

Moran, S., A. Martinez-Cardus, S. Sayols, E. Musulen, C. Balana, A. Estival-Gonzalez, C. Moutinho, H. Heyn, A. Diaz-Lagares, M. C. de Moura, G. M. Stella, P. M. Comoglio, M. Ruiz-Miro, X. Matias-Guiu, R. Pazo-Cid, A. Anton, R. Lopez-Lopez, G. Soler, F. Longo, I. Guerra, S. Fernandez, Y. Assenov, C. Plass, R. Morales, J. Carles, D. Bowtell, L. Mileshkin, D. Sia, R. Tothill, J. Tabernero, J. M. Llovet, and M. Esteller. 2016. 'Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis', *Lancet Oncol*, 17: 1386-95.

Mueller, W. C., Y. Spector, T. B. Edmonston, B. St Cyr, D. Jaeger, U. Lass, R. Aharonov, S. Rosenwald, and A. Chajut. 2011. 'Accurate classification of metastatic brain tumors using a novel microRNA-based test', *Oncologist*, 16: 165-74.

NCCN. v.2/2023. 'NCCN guidelines for occult primary tumors'.

Nystrom, J. S., J. M. Weiner, J. Heffelfinger-Juttner, L. E. Irwin, J. R. Bateman, and R. M. Wolf. 1977. 'Metastatic and histologic presentations in unknown primary cancer', *Semin Oncol*, 4: 53-8.

Oien, K. A. 2009. 'Pathologic evaluation of unknown primary cancer', *Semin Oncol*, 36: 8-37.

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

Pavlidis, N., and G. Pentheroudakis. 2012. 'Cancer of unknown primary site', *Lancet*, 379: 1428-35.

Peiro-Chova, L., M. Pena-Chilet, J. A. Lopez-Guerrero, J. L. Garcia-Gimenez, E. Alonso-Yuste, O. Burgues, A. Lluch, J. Ferrer-Lozano, and G. Ribas. 2013. 'High stability of microRNAs in tissue samples of compromised quality', *Virchows Arch*, 463: 765-74.

Pentheroudakis, G., E. Briasoulis, H. P. Kalofonos, G. Fountzilas, T. Economopoulos, G. Samelis, A. Koutras, M. Karina, N. Xiros, E. Samantas, A. Bamias, N. Pavlidis, and Group Hellenic Cooperative Oncology. 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: 1148-55.

Pentheroudakis, G., V. Golfinopoulos, and N. Pavlidis. 2007. 'Switching benchmarks in cancer of unknown primary: from autopsy to microarray', *Eur J Cancer*, 43: 2026-36.

Pentheroudakis, G., E. A. Kotteas, V. Kotoula, K. Papadopoulou, E. Charalambous, A. Cervantes, T. Ciuleanu, G. Fountzilas, and N. Pavlidis. 2014. 'Mutational profiling of the RAS, PI3K, MET and b-catenin pathways in cancer of unknown primary: a retrospective study of the Hellenic Cooperative Oncology Group', *Clin Exp Metastasis*, 31: 761-9.

Pentheroudakis, G., N. Pavlidis, G. Fountzilas, D. Krikelis, A. Goussia, A. Stoyianni, M. Sanden, B. St Cyr, N. Yerushalmi, H. Benjamin, E. Meiri, A. Chajut, S. Rosenwald, R. Aharonov, and Y. Spector. 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.

Piana, A. F., G. Sotgiu, M. R. Muroni, P. Cossu-Rocca, P. Castiglia, and M. R. De Miglio. 2014. 'HPV infection and triple-negative breast cancers: an Italian case-control study', *Virol J*, 11: 190.

Pillai, R., R. Deeter, C. T. Rigl, J. S. Nystrom, M. H. Miller, L. Buturovic, and W. D. Henner. 2011. 'Validation and reproducibility of a microarray-based gene expression test for tumor identification in formalin-fixed, paraffin-embedded specimens', *J Mol Diagn*, 13: 48-56.

Ramaswamy, S., P. Tamayo, R. Rifkin, S. Mukherjee, C. H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, and T. R. Golub. 2001.

'Multiclass cancer diagnosis using tumor gene expression signatures', *Proc Natl Acad Sci U S A*, 98: 15149-54.

Ramskold, D., E. T. Wang, C. B. Burge, and R. Sandberg. 2009. 'An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data', *PLoS Comput Biol*, 5: e1000598.

Rosenfeld, N., R. Aharonov, E. Meiri, S. Rosenwald, Y. Spector, M. Zepeniuk, H. Benjamin, N. Shabes, S. Tabak, A. Levy, D. Lebanony, Y. Goren, E. Silberschein, N. Targan, A. Ben-Ari, S. Gilad, N. Sion-Vardy, A. Tobar, M. Feinmesser, O. Kharenko, O. Nativ, D. Nass, M. Perelman, A. Yosepovich, B. Shalmon, S. Polak-Charcon, E. Fridman, A. Avniel, I. Bentwich, Z. Bentwich, D. Cohen, A. Chajut, and I. Barshack. 2008. 'MicroRNAs accurately identify cancer tissue origin', *Nat Biotechnol*, 26: 462-9.

Rosenwald, S., S. Gilad, S. Benjamin, D. Lebanony, N. Dromi, A. Faerman, H. Benjamin, R. Tamir, M. Ezagouri, E. Goren, I. Barshack, D. Nass, A. Tobar, M. Feinmesser, N. Rosenfeld, I. Leizerman, K. Ashkenazi, Y. Spector, A. Chajut, and R. Aharonov. 2010. 'Validation of a microRNA-based qRT-PCR test for accurate identification of tumor tissue origin', *Mod Pathol*, 23: 814-23.

Ross, J. S., K. Wang, L. Gay, G. A. Otto, E. White, K. Iwanik, G. Palmer, R. Yelensky, D. M. Lipson, J. Chmielecki, R. L. Erlich, A. N. Rankin, S. M. Ali, J. A. Elvin, D. Morosini, V. A. Miller, and P. J. Stephens. 2015. 'Comprehensive Genomic Profiling of Carcinoma of Unknown Primary Site: New Routes to Targeted Therapies', *JAMA Oncol*, 1: 40-49.

Schweizer, J., P. E. Bowden, P. A. Coulombe, L. Langbein, E. B. Lane, T. M. Magin, L. Maltais, M. B. Omary, D. A. Parry, M. A. Rogers, and M. W. Wright. 2006. 'New consensus nomenclature for mammalian keratins', *J Cell Biol*, 174: 169-74.

Shimonishi, T., K. Miyazaki, and Y. Nakanuma. 2000. 'Cytokeratin profile relates to histological subtypes and intrahepatic location of intrahepatic cholangiocarcinoma and primary sites of metastatic adenocarcinoma of liver', *Histopathology*, 37: 55-63.

Smyth, E. C., I. S. Babina, and N. C. Turner. 2017. 'Gatekeeper Mutations and Intratumoral Heterogeneity in FGFR2-Translocated Cholangiocarcinoma', *Cancer Discov*, 7: 248-49.

Snyder, R. D., G. M. Mavligit, and M. Valdivieso. 1979. 'Adenocarcinoma of unknown primary site: a clinico-pathological study', *Med Pediatr Oncol*, 6: 389-94.

Sokilde, R., M. Vincent, A. K. Moller, A. Hansen, P. E. Hoiby, T. Blondal, B. S. Nielsen, G. Daugaard, S. Moller, and T. Litman. 2014. 'Efficient identification of miRNAs for classification of tumor origin', *J Mol Diagn*, 16: 106-15.

Stewart, J. F., M. H. Tattersall, R. L. Woods, and R. M. Fox. 1979. 'Unknown primary adenocarcinoma: incidence of overinvestigation and natural history', *Br Med J*, 1: 1530-3.

Su, A. I., J. B. Welsh, L. M. Sapinoso, S. G. Kern, P. Dimitrov, H. Lapp, P. G. Schultz, S. M. Powell, C. A. Moskaluk, H. F. Frierson, Jr., and G. M. Hampton. 2001. 'Molecular classification of human carcinomas by use of gene expression signatures', *Cancer Res*, 61: 7388-93.

Subbiah, I. M., A. Tsimberidou, V. Subbiah, F. Janku, S. Roy-Chowdhuri, and D. S. Hong. 2017. 'Next generation sequencing of carcinoma of unknown primary reveals novel combinatorial strategies in a heterogeneous mutational landscape', *Oncoscience*, 4: 47-56.

Talantov, D., J. Baden, T. Jatkoe, K. Hahn, J. Yu, Y. Rajpurohit, Y. Jiang, C. Choi, J. S. Ross, D. Atkins, Y. Wang, and A. Mazumder. 2006. 'A quantitative reverse transcriptase-polymerase chain reaction assay to identify metastatic carcinoma tissue of origin', *J Mol Diagn*, 8: 320-9.

Tang, W., S. Wan, Z. Yang, A. E. Teschendorff, and Q. Zou. 2018. 'Tumor origin detection with tissue-specific miRNA and DNA methylation markers', *Bioinformatics*, 34: 398-406.

Tibshirani, R. J. 2009. 'Univariate shrinkage in the cox model for high dimensional data', *Stat Appl Genet Mol Biol*, 8: Article21.

Tibshirani, Robert. 1996. 'Regression Shrinkage and Selection Via the Lasso', *Journal of the Royal Statistical Society: Series B (Methodological)*, 58: 267-88.

Tothill, R. W., A. Kowalczyk, D. Rischin, A. Bousioutas, I. Haviv, R. K. van Laar, P. M. Waring, J. Zalcberg, R. Ward, A. V. Biankin, R. L. Sutherland, S. M. Henshall, K. Fong, J. R. Pollack, D. D. Bowtell, and A. J. Holloway. 2005. 'An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin', *Cancer Res*, 65: 4031-40.

Tothill, R. W., J. Li, L. Mileshkin, K. Doig, T. Siganakis, P. Cowin, A. Fellowes, T. Semple, S. Fox, K. Byron, A. Kowalczyk, D. Thomas, P. Schofield, and D. D. Bowtell. 2013. 'Massively-parallel sequencing assists the diagnosis and guided treatment of cancers of unknown primary', *J Pathol*, 231: 413-23.

Tothill, R. W., F. Shi, L. Paiman, J. Bedo, A. Kowalczyk, L. Mileshkin, E. Buela, R. Klupacs, D. Bowtell, and K. Byron. 2015. 'Development and validation of a gene expression tumour classifier for cancer of unknown primary', *Pathology*, 47: 7-12.

Tsimberidou, A. M., D. S. Hong, Y. Ye, C. Cartwright, J. J. Wheler, G. S. Falchook, A. Naing, S. Fu, S. Piha-Paul, F. Janku, F. Meric-Bernstam, P. Hwu, B. Kee, M. S. Kies, R. Broaddus, J. Mendelsohn, K. R. Hess, and R. Kurzrock. 2017. 'Initiative for Molecular Profiling and Advanced Cancer Therapy (IMPACT): An MD Anderson Precision Medicine Study', *JCO Precis Oncol*, 2017.

Urban, D., A. Rao, M. Bressel, Y. R. Lawrence, and L. Mileshkin. 2013. 'Cancer of unknown primary: a population-based analysis of temporal change and socioeconomic disparities', *Br J Cancer*, 109: 1318-24.

Vaideeswar, P., S. Patil, and J. Chaudhari. 2022. 'Disseminated epithelial cancers-An autopsy analysis', *Indian J Pathol Microbiol*, 65: 76-86.

Varadhachary, G. R., and M. N. Raber. 2014. 'Carcinoma of unknown primary site', *N Engl J Med*, 371: 2040.

Varadhachary, G. R., Y. Spector, J. L. Abbruzzese, S. Rosenwald, H. Wang, R. Aharonov, H. R. Carlson, D. Cohen, S. Karanth, J. Macinskas, R. Lenzi, A. Chajut, T. B. Edmonston, and M. N. Raber. 2011. 'Prospective gene signature study using microRNA to identify the tissue of origin in patients with carcinoma of unknown primary', *Clin Cancer Res*, 17: 4063-70.

Varadhachary, G. R., D. Talantov, M. N. Raber, C. Meng, K. R. Hess, T. Jatkoe, R. Lenzi, D. R. Spigel, Y. Wang, F. A. Greco, J. L. Abbruzzese, and J. D. Hainsworth. 2008. 'Molecular profiling of carcinoma of unknown primary and correlation with clinical evaluation', *J Clin Oncol*, 26: 4442-8.

Varghese, A. M., A. Arora, M. Capanu, N. Camacho, H. H. Won, A. Zehir, J. Gao, D. Chakravarty, N. Schultz, D. S. Klimstra, M. Ladanyi, D. M. Hyman, D. B. Solit, M. F. Berger, and L. B. Saltz. 2017. 'Clinical and molecular characterization of patients with cancer of unknown primary in the modern era', *Ann Oncol*, 28: 3015-21.

Verginelli, F., A. Pisacane, G. Gambardella, A. D'Ambrosio, E. Candiello, M. Ferrio, M. Panero, L. Casorzo, S. Benvenuti, E. Cascardi, R. Senetta, E. Geuna, A. Ballabio, F. Montemurro, A.

Sapino, P. M. Comoglio, and C. Boccaccio. 2021. 'Cancer of unknown primary stem-like cells model multi-organ metastasis and unveil liability to MEK inhibition', *Nat Commun*, 12: 2498.

von Ahlfen, S., A. Missel, K. Bendrat, and M. Schlumpberger. 2007. 'Determinants of RNA quality from FFPE samples', *PLoS One*, 2: e1261.

Wainberg, Z. A., P. C. Enzinger, Y. K. Kang, S. Qin, K. Yamaguchi, I. H. Kim, A. Saeed, S. C. Oh, J. Li, H. M. Turk, A. Teixeira, C. Borg, E. Hitre, A. A. Udrea, G. G. Cardellino, R. G. Sanchez, H. Collins, S. Mitra, Y. Yang, D. V. T. Catenacci, and K. W. Lee. 2022. 'Bemarituzumab in patients with FGFR2b-selected gastric or gastro-oesophageal junction adenocarcinoma (FIGHT): a randomised, double-blind, placebo-controlled, phase 2 study', *Lancet Oncol*, 23: 1430-40.

Wang, K., M. Li, and H. Hakonarson. 2010. 'ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data', *Nucleic Acids Res*, 38: e164.

Warner, E., R. Goel, J. Chang, W. Chow, S. Verma, J. Dancey, E. Franssen, H. Dulude, M. Girouard, J. Correia, and G. Gallant. 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: 2376-80.

Weiss, L. M., P. Chu, B. E. Schroeder, V. Singh, Y. Zhang, M. G. Erlander, and C. A. Schnabel. 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: 263-9.

Wildi, S., J. Kleeff, H. Maruyama, C. A. Maurer, H. Friess, M. W. Buchler, A. D. Lander, and M. Korc. 1999. 'Characterization of cytokeratin 20 expression in pancreatic and colorectal cancer', *Clin Cancer Res*, 5: 2840-7.

Yoon, H. H., N. R. Foster, J. P. Meyers, P. D. Steen, D. W. Visscher, R. Pillai, D. M. Prow, C. M. Reynolds, B. T. Marchello, R. B. Mowat, B. I. Mattar, C. Erlichman, and M. P. Goetz. 2016. 'Gene expression profiling identifies responsive patients with cancer of unknown primary treated with carboplatin, paclitaxel, and everolimus: NCCTG N0871 (alliance)', *Ann Oncol*, 27: 339-44.

Zhang, Y., R. Wang, Y. Li, Y. Pan, H. Hu, Y. Zhang, H. Li, L. Shen, Y. Yu, Y. Sun, and H. Chen. 2015. 'Negative Thyroid Transcription Factor 1 Expression Defines an Unfavorable Subgroup of Lung Adenocarcinomas', *J Thorac Oncol*, 10: 1444-50.

7 SUPPLEMENTARY TABLES

Supplementary Table 1. Average miRNA expression and standard deviations for each tumor class.

	LBC		TNBC		CHOL		CRC		UCEC		STAD		GI-NET	
	mean	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**
let-7e-5p	22,760	15,150	4,533	1,464	3,750	3,294	4,636	5,414	29,918	42,305	12,003	13,045	56,742	77,152
let-7i-5p	84,300	84,972	51,367	9,411	23,850	23,391	52,233	36,928	302,800	283,656	43,442	36,429	131,660	154,150
miR-10a-5p	2,468	1,417	0,730	0,619	11,507	17,108	11,386	6,544	51,380	65,520	3,474	2,756	14,184	19,684
miR-10a-3p	0,094	0,104	0,237	0,255	0,298	0,323	0,423	0,256	0,834	0,483	0,238	0,297	0,270	0,521
miR-10b-5p	15,340	7,976	2,867	1,343	1,643	1,714	7,257	3,945	187,500	80,310	6,080	3,687	4,652	4,480
miR-106b-5p	17,120	15,196	25,400	13,289	12,150	11,539	16,400	14,839	76,760	92,245	86,780	169,068	14,580	16,024
miR-122-5p	0,010	0,000	0,010	0,000	432,000	727,567	0,190	0,402	1,110	1,983	0,828	1,453	0,040	0,067
miR-124-3p	0,086	0,082	0,030	0,035	0,043	0,052	0,154	0,105	1,120	1,451	0,054	0,074	0,276	0,411
miR-126-5p	3,074	4,195	0,387	0,326	2,290	3,018	0,460	0,999	21,360	25,699	1,444	2,779	3,482	3,334
miR-130a-3p	19,660	20,120	20,633	6,519	6,100	3,416	5,557	4,166	256,220	383,123	10,520	10,078	10,760	11,171
miR-135b-5p	10,596	21,592	19,867	20,709	13,672	20,626	20,329	18,005	61,520	18,237	7,000	7,814	1,244	0,709
miR-138-5p	0,652	1,214	0,030	0,035	0,027	0,041	0,124	0,250	0,428	0,598	0,274	0,226	3,150	6,101
miR-141-3p	51,020	31,516	114,133	59,166	13,338	31,044	66,957	61,292	177,040	140,795	43,620	40,950	125,600	135,461
miR-142-3p	54,160	60,291	167,333	14,012	49,783	88,577	52,714	61,960	200,700	163,454	16,560	13,820	20,020	22,725
miR-145-5p	190,800	99,115	55,100	12,853	368,883	438,101	222,286	111,412	2842,260	3052,932	334,000	357,956	998,260	1603,129
miR-146a-5p	18,080	21,523	13,300	1,873	9,867	7,490	11,743	6,872	34,240	17,559	12,980	11,676	7,978	5,805
miR-148b-3p	4,188	3,959	3,300	0,265	2,100	1,562	1,607	1,623	23,420	23,191	4,186	3,209	23,030	29,476
miR-149-5p	5,960	3,870	0,687	0,241	1,302	1,582	0,807	0,629	7,300	9,357	1,414	1,730	8,040	13,711
miR-152-3p	10,620	2,616	6,700	1,562	14,203	13,401	6,357	5,328	60,120	84,456	4,460	2,273	14,760	10,080
miR-181a-5p	36,140	17,475	20,433	5,160	21,150	16,180	19,086	16,074	71,640	57,406	16,280	11,227	15,340	10,288
miR-181a-2-3p	0,546	0,445	0,300	0,106	1,122	0,853	0,371	0,233	1,806	0,822	0,274	0,142	0,380	0,281
miR-181b-5p	15,940	11,214	9,367	2,359	9,700	6,587	6,157	3,073	32,900	16,070	6,460	4,746	7,840	4,422
miR-182-5p	12,400	6,474	2,033	0,924	3,082	2,088	2,657	2,213	8,320	2,536	2,378	1,204	17,124	23,140
miR-183-5p	5,775	2,724	1,620	1,022	2,442	2,245	2,592	1,572	6,060	3,548	1,678	1,072	23,950	31,596

miR-187-3p	3,080	2,805	0,260	0,382	0,848	0,976	0,793	0,909	3,200	1,288	1,236	1,192	4,206	6,465
miR-187-5p	0,010	0,000	0,037	0,046	0,018	0,020	0,010	0,000	0,044	0,053	0,010	0,000	0,082	0,135
miR-192-5p	1,356	1,471	0,523	0,260	66,083	58,519	98,286	55,374	22,120	25,567	93,320	31,314	238,860	307,248
miR-193a-3p	3,820	2,170	2,733	1,531	0,818	1,191	3,160	3,356	7,328	11,018	2,090	2,473	2,470	2,584
miR-193b-3p	24,420	12,248	12,200	7,351	11,783	7,455	10,343	4,328	60,360	83,744	10,036	11,320	10,960	5,613
miR-194-5p	2,644	2,263	1,573	0,766	136,883	140,548	161,543	76,795	48,200	46,705	185,400	65,926	412,220	563,742
miR-194-3p	0,988	1,763	0,077	0,115	0,750	1,305	2,946	3,794	0,252	0,253	1,678	2,559	2,070	1,826
miR-196a-5p	28,460	17,788	3,933	2,550	1,010	1,393	5,614	4,709	3,834	2,900	5,380	2,377	29,848	65,491
miR-19b-3p	53,620	53,914	104,667	23,180	58,267	48,069	90,671	72,163	488,080	822,011	34,260	20,843	45,880	34,180
miR-200a-3p	11,760	16,222	9,700	5,110	5,834	10,678	10,714	4,284	29,900	6,690	19,040	14,975	105,200	135,190
miR-200a-5p	0,406	0,198	0,190	0,197	0,388	0,471	0,780	0,795	1,796	0,787	0,886	0,439	1,992	2,551
miR-200b-3p	49,980	32,375	38,367	4,966	54,517	55,194	126,686	55,713	231,200	62,588	94,220	64,074	291,500	390,035
miR-200b-5p	1,842	0,685	1,900	0,900	1,845	1,575	4,129	2,866	4,660	1,137	3,160	3,349	5,812	6,585
miR-200c-3p	123,000	89,035	94,267	44,802	45,628	106,495	144,686	46,076	174,700	52,043	108,660	70,762	332,220	305,140
miR-204-5p	2,502	1,526	0,473	0,411	2,687	3,273	2,090	4,771	35,320	17,001	1,036	1,056	13,220	22,288
miR-205-5p	180,000	92,978	259,667	210,438	1,210	1,988	2,336	3,376	80,820	112,752	3,212	5,139	0,824	0,788
miR-210-3p	8,274	7,161	13,033	3,761	1,942	2,008	11,300	8,445	20,200	12,015	8,260	4,792	7,626	9,439
miR-211-5p	3,802	3,676	0,027	0,029	0,434	0,640	2,143	4,067	2,010	3,696	2,378	4,118	2,064	2,736
miR-214-3p	32,220	10,252	9,100	4,321	28,783	23,783	14,814	11,260	149,020	196,600	19,240	22,472	43,720	60,862
miR-215-5p	1,226	1,201	0,717	0,333	64,467	63,732	91,157	47,647	21,660	23,039	98,840	32,626	207,360	269,531
miR-27b-3p	100,780	84,117	3,870	4,840	10,085	9,425	20,443	22,358	237,960	189,821	38,360	40,736	248,660	329,060
miR-29a-3p	79,420	75,568	67,633	26,461	49,417	39,667	49,967	33,500	653,000	802,428	36,640	29,878	96,680	117,298
miR-29b-3p	43,240	26,199	52,267	19,299	18,167	16,273	44,500	35,859	533,460	970,023	23,540	14,183	118,000	143,875
miR-29c-3p	119,780	69,152	118,000	22,068	90,733	70,973	87,014	57,077	1557,400	2743,883	57,340	51,853	233,200	250,682
miR-30a-5p	29,380	23,264	3,200	3,985	12,950	3,835	2,349	1,630	28,940	17,574	4,744	5,122	19,360	17,625
miR-30a-3p	6,360	4,044	1,127	1,078	5,933	2,926	0,689	1,033	12,320	14,228	1,472	1,129	4,080	2,973
miR-30c-5p	75,800	36,622	47,367	9,903	177,217	95,173	23,543	9,101	446,600	524,217	29,500	21,643	119,360	86,744
miR-31-5p	5,560	5,187	7,380	10,680	6,965	15,598	3,627	4,364	28,180	10,194	18,160	15,061	3,094	4,417
miR-31-3p	1,870	2,411	0,247	0,270	0,377	0,702	0,857	0,881	2,660	1,332	1,200	0,861	0,256	0,264
miR-340-3p	2,812	1,698	0,273	0,232	0,657	0,475	1,899	2,061	1,530	1,186	2,156	2,459	1,090	0,836
miR-342-3p	54,760	18,837	16,167	3,702	28,717	27,170	7,471	5,929	90,480	83,010	15,280	19,274	45,220	57,760

miR-345-5p	7,720	7,028	4,133	0,404	1,585	1,574	5,571	4,307	25,620	16,699	5,260	1,491	9,500	10,843
miR-34a-5p	34,020	24,659	24,267	11,673	8,233	9,346	15,971	9,294	123,820	85,842	12,700	8,059	19,340	26,466
miR-34b-3p	0,138	0,150	0,010	0,000	0,140	0,231	0,045	0,086	2,412	3,769	0,036	0,037	0,678	1,410
miR-361-5p	16,560	9,733	10,467	3,259	14,567	10,192	6,571	4,121	127,180	163,736	8,280	3,871	34,760	38,019
miR-363-3p	4,538	4,092	0,997	0,463	2,708	3,388	0,849	0,981	7,260	5,363	1,832	0,945	1,670	1,101
miR-372-3p	0,900	1,107	0,050	0,036	0,420	0,589	0,018	0,020	0,084	0,103	7,906	17,053	0,314	0,556
miR-373-3p	0,472	0,528	0,010	0,000	0,172	0,257	0,414	0,964	0,186	0,129	13,608	30,294	0,290	0,465
miR-375	29,225	46,547	2,037	1,879	10,962	13,154	35,357	35,389	29,000	11,408	112,000	224,860	2784,800	3482,900
miR-382-5p	0,888	0,300	0,383	0,107	1,002	0,929	1,376	1,002	1,512	1,176	0,578	0,314	0,876	0,846
miR-485-5p	0,144	0,158	0,010	0,000	0,073	0,105	0,010	0,000	0,090	0,076	0,146	0,304	0,124	0,124
miR-506-3p	0,968	0,958	0,010	0,000	0,140	0,219	1,106	1,744	0,654	0,937	0,826	1,109	0,590	0,346
miR-508-3p	2,242	2,860	0,010	0,000	0,120	0,246	0,583	1,127	0,568	0,549	0,960	1,091	0,390	0,353
miR-509-3p	0,758	0,683	0,030	0,035	0,352	0,617	1,346	2,235	0,404	0,055	2,054	2,405	2,654	5,016
miR-510-5p	0,208	0,443	0,010	0,000	0,145	0,331	0,010	0,000	0,010	0,000	0,124	0,158	0,328	0,711
miR-514a-3p	0,566	0,826	0,123	0,196	0,010	0,000	0,475	1,139	0,068	0,130	0,530	0,634	0,328	0,711
miR-552-3p	0,054	0,070	0,010	0,000	0,053	0,080	0,719	0,354	0,098	0,121	0,310	0,430	0,250	0,367
miR-649	0,962	1,291	0,040	0,052	0,078	0,140	0,269	0,506	0,130	0,268	0,114	0,233	0,056	0,063
miR-650	0,068	0,070	0,010	0,000	0,020	0,024	0,141	0,147	0,138	0,286	0,214	0,296	0,062	0,116
miR-661	1,016	0,758	0,010	0,000	0,317	0,551	0,584	0,952	0,320	0,437	1,152	1,238	1,126	0,805
miR-873-5p	0,010	0,000	0,110	0,173	0,010	0,000	0,037	0,034	3,074	5,508	0,010	0,000	0,194	0,411
miR-9-3p	0,086	0,144	1,033	1,531	0,235	0,265	0,201	0,335	5,966	10,234	0,010	0,000	0,110	0,137
miR-92a-3p	86,260	108,031	77,667	26,035	55,483	26,963	102,157	50,540	260,240	269,393	88,440	67,947	59,580	44,409
miR-92b-3p	5,500	2,830	1,867	0,321	0,792	0,577	1,563	1,148	15,960	13,678	1,930	1,538	6,726	10,672
miR-96-5p	7,820	5,374	3,600	0,866	0,798	0,978	3,071	2,492	11,860	5,770	3,190	3,559	20,488	27,663
miR-99a-5p	149,840	170,270	52,633	25,434	55,550	50,397	17,100	9,666	2126,800	2840,054	38,560	28,191	101,460	114,697
miR-24-3p	166,400	80,039	122,000	13,229	315,050	267,440	120,629	85,103	1660,000	2228,519	128,220	121,209	351,240	306,665
miR-21-5p	870,200	689,090	330,333	334,434	2826,000	5911,414	691,571	495,020	1417,000	663,586	472,600	290,887	444,800	333,854
miR-16-5p	247,800	123,749	157,000	19,519	251,833	102,355	80,371	45,570	644,600	719,485	97,400	54,917	242,400	220,416
miR-320a	39,700	33,963	20,600	1,646	38,333	31,035	43,243	19,804	196,360	211,834	49,220	17,021	47,940	45,644
miR-224-5p	2,768	4,570	2,097	2,777	0,560	0,314	4,471	3,517	2,400	1,075	1,084	1,013	1,062	1,737
miR-423-5p	4,944	5,639	1,087	1,196	0,592	0,690	2,696	2,187	13,720	10,246	3,082	4,044	7,244	9,227

miR-25-3p	46,860	32,297	49,467	20,835	59,167	24,851	44,486	23,580	132,880	152,132	50,020	21,686	36,740	21,673
miR-331-3p	12,860	5,723	6,400	2,390	4,617	4,013	6,614	2,795	45,360	53,358	8,480	5,081	25,638	29,847
miR-103a-3p	87,002	94,094	59,400	13,455	29,150	25,909	31,843	22,255	282,200	243,005	53,700	21,043	175,040	198,550

	HNSC		KIRC		KIRP		LIHC		LUAD		LUSC		OV	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
let-7e-5p	1,387	1,531	12,980	9,537	7,400	1,697	0,870	1,290	10,460	11,674	9,267	5,980	15,180	17,323
let-7i-5p	47,767	32,365	98,420	77,683	74,933	39,926	36,100	35,449	148,683	122,443	140,300	70,173	265,733	452,704
miR-10a-5p	0,702	1,333	6,048	6,572	66,033	60,143	2,335	2,520	7,230	7,847	16,933	10,366	78,557	182,006
miR-10a-3p	0,020	0,024	0,224	0,196	0,973	0,575	0,108	0,060	0,193	0,218	0,187	0,095	2,143	4,977
miR-10b-5p	1,628	0,967	48,920	37,993	46,567	52,501	1,773	1,355	2,537	2,649	11,767	6,574	163,183	217,292
miR-106b-5p	15,883	11,332	23,200	10,577	20,267	27,742	8,917	10,252	24,583	17,177	73,500	59,326	182,250	344,240
miR-122-5p	0,022	0,029	2,562	2,791	0,573	0,976	1708,683	1970,098	0,022	0,029	0,010	0,000	0,375	0,655
miR-124-3p	0,070	0,097	0,310	0,304	0,067	0,098	0,010	0,000	0,200	0,295	0,090	0,139	0,213	0,274
miR-126-5p	0,062	0,127	40,496	72,252	6,570	10,332	79,250	186,093	8,733	12,878	2,137	3,105	4,045	4,461
miR-130a-3p	5,400	2,529	20,700	13,896	28,600	24,694	11,000	12,133	22,900	20,591	23,900	13,235	250,267	392,263
miR-135b-5p	6,300	4,042	2,190	2,748	25,587	40,928	0,237	0,345	47,400	63,080	58,067	40,921	462,800	1008,212
miR-138-5p	0,010	0,000	0,278	0,251	1,770	2,174	0,010	0,000	0,888	1,004	0,533	0,497	0,092	0,200
miR-141-3p	71,617	61,520	0,322	0,180	71,173	75,158	0,525	0,836	222,617	207,480	140,667	48,881	277,833	241,087
miR-142-3p	164,433	109,149	79,140	72,108	263,967	252,552	271,000	530,265	266,983	256,734	154,000	70,711	187,583	196,024
miR-145-5p	114,700	118,210	344,000	218,410	99,367	157,377	120,817	152,021	268,083	209,504	471,000	124,964	350,450	335,202
miR-146a-5p	8,083	8,047	22,240	13,356	3,033	4,415	6,700	7,902	66,767	70,048	28,267	17,679	18,690	13,897
miR-148b-3p	2,010	1,300	7,000	5,331	5,333	5,992	4,168	6,627	7,583	9,023	10,133	2,250	7,370	7,097
miR-149-5p	6,347	6,136	1,726	2,676	6,300	7,238	0,785	1,411	3,447	5,416	20,433	9,857	13,422	15,158
miR-152-3p	8,367	6,642	11,880	5,573	23,767	18,536	97,933	192,753	12,167	14,688	36,100	16,813	19,900	14,875
miR-181a-5p	9,667	4,745	37,300	17,969	19,600	30,148	16,917	17,939	107,017	85,357	66,767	16,954	34,017	22,845
miR-181a-2-3p	0,233	0,090	0,638	0,319	3,903	5,504	0,682	0,971	1,203	0,560	1,450	1,191	1,025	0,388
miR-181b-5p	5,567	2,980	19,080	8,281	372,977	623,677	8,157	10,640	38,150	35,543	26,400	5,237	15,717	7,356
miR-182-5p	4,243	5,384	0,840	0,734	7,033	4,086	15,968	33,433	8,600	7,533	13,167	5,390	5,700	5,059
miR-183-5p	4,253	6,170	1,812	1,549	3,867	2,401	8,962	14,740	6,590	6,565	10,800	6,151	4,650	4,387
miR-187-3p	3,087	4,771	0,678	0,561	0,297	0,437	0,678	1,067	2,720	2,741	6,600	2,905	5,900	6,410
miR-187-5p	0,023	0,033	0,010	0,000	2,207	3,805	0,036	0,058	0,010	0,000	0,027	0,029	0,140	0,240
miR-192-5p	0,760	0,701	49,380	33,067	3,383	5,731	1155,100	2402,250	2,108	1,501	4,383	4,984	282,367	543,554
miR-193a-3p	1,345	1,114	3,812	4,154	31,037	52,796	2,603	2,878	2,002	1,742	2,033	0,231	2,820	2,818
miR-193b-3p	16,867	12,712	11,420	12,195	9,500	7,312	38,233	16,970	11,883	6,766	38,600	15,120	46,000	59,084

miR-194-5p	1,255	0,785	86,160	56,827	7,287	11,278	1036,333	1720,164	3,230	2,944	10,667	13,310	430,500	808,427
miR-194-3p	0,032	0,053	0,214	0,200	5,360	9,215	1,512	2,185	0,062	0,066	0,027	0,029	1,990	3,017
miR-196a-5p	2,637	5,197	22,980	15,578	62,857	93,245	0,483	0,850	3,662	4,963	2,610	3,727	5,827	4,572
miR-19b-3p	60,433	33,112	59,900	32,477	200,250	216,021	85,567	120,607	86,467	61,223	158,667	62,804	203,200	174,963
miR-200a-3p	1,823	0,948	8,368	5,802	29,737	45,239	0,305	0,236	14,718	19,766	13,533	5,150	119,067	212,653
miR-200a-5p	0,265	0,282	0,204	0,165	6,117	9,108	0,097	0,114	0,553	0,481	1,303	1,142	9,887	21,520
miR-200b-3p	35,400	27,887	37,720	26,369	146,480	226,189	9,542	14,210	81,917	101,561	155,033	95,101	1158,517	2402,773
miR-200b-5p	1,425	0,731	0,780	0,581	3,567	3,553	0,352	0,375	2,027	1,965	3,900	2,884	67,683	150,113
miR-200c-3p	47,720	30,443	0,914	0,530	26,183	33,061	2,917	6,754	352,500	329,154	299,000	102,328	1001,333	1589,971
miR-204-5p	0,280	0,552	146,280	139,265	37,383	64,620	3,650	5,763	0,607	0,574	1,557	1,169	20,047	39,266
miR-205-5p	433,167	311,615	2,558	3,652	0,520	0,765	1,503	2,474	228,067	314,282	3203,333	502,129	399,767	932,040
miR-210-3p	8,745	10,403	78,000	45,202	11,937	15,532	5,767	5,011	22,400	18,044	77,000	23,587	208,467	444,724
miR-211-5p	0,242	0,567	0,988	1,852	6,260	10,773	1,655	2,683	1,108	1,443	0,340	0,572	0,522	0,975
miR-214-3p	12,433	15,172	6,020	2,180	9,713	11,113	10,477	12,611	20,950	13,928	32,000	10,701	127,967	156,882
miR-215-5p	0,475	0,219	42,100	29,054	4,680	5,968	1047,233	2129,025	2,022	1,510	5,230	6,704	115,767	130,474
miR-27b-3p	4,142	3,670	80,840	103,498	184,067	186,724	332,285	790,589	22,197	34,976	129,333	56,039	113,117	177,468
miR-29a-3p	31,100	20,251	62,920	37,331	262,667	73,657	69,217	87,802	110,650	79,776	65,100	9,814	321,600	349,910
miR-29b-3p	47,133	47,311	54,160	47,058	346,333	276,560	70,333	91,165	96,833	88,669	46,867	11,676	403,317	460,221
miR-29c-3p	97,500	88,850	123,640	100,032	372,167	311,342	150,317	215,900	268,717	227,980	164,333	53,163	474,500	539,515
miR-30a-5p	1,648	1,212	66,720	39,721	32,933	18,513	9,922	15,288	9,217	5,037	10,567	5,925	38,017	25,518
miR-30a-3p	0,765	0,585	21,620	11,715	192,500	283,663	7,613	14,439	4,812	6,848	3,100	1,217	11,200	13,378
miR-30c-5p	33,450	13,338	295,400	163,598	413,867	591,028	407,533	758,730	105,850	106,999	398,667	487,155	383,667	281,697
miR-31-5p	3,588	4,500	2,312	2,375	58,850	100,591	0,697	1,377	9,103	8,292	158,667	68,413	293,703	625,343
miR-31-3p	0,075	0,071	0,398	0,236	4,743	7,508	0,173	0,268	0,880	0,886	10,867	7,808	19,098	40,708
miR-340-3p	0,192	0,231	1,518	1,838	6,183	10,067	4,628	8,945	0,532	0,514	0,747	0,356	0,657	0,362
miR-342-3p	9,217	6,039	34,480	22,033	17,067	22,027	47,800	64,030	36,717	29,145	50,133	10,306	65,783	38,966
miR-345-5p	3,183	1,646	7,380	4,011	9,637	8,748	7,483	4,859	16,517	15,414	16,067	6,513	36,550	29,967
miR-34a-5p	24,400	23,576	58,320	51,890	62,537	97,599	28,050	26,399	52,500	50,358	45,633	20,775	71,980	76,594
miR-34b-3p	0,093	0,128	0,086	0,074	9,023	15,568	0,020	0,024	0,393	0,408	0,803	0,390	0,530	0,874
miR-361-5p	7,317	4,887	21,980	10,179	16,933	21,084	32,517	46,845	22,067	19,769	26,533	9,790	42,250	13,389
miR-363-3p	1,403	1,342	5,140	3,428	2,203	2,758	2,513	2,025	3,302	2,316	3,467	1,762	5,050	2,005

miR-372-3p	0,018	0,020	0,428	0,935	0,010	0,000	0,682	1,090	0,025	0,037	0,030	0,035	0,273	0,455
miR-373-3p	0,065	0,074	0,712	0,899	0,060	0,087	0,543	0,885	0,348	0,494	0,133	0,143	0,123	0,091
miR-375	3,145	4,381	0,068	0,068	0,520	0,596	23,523	53,666	50,383	72,452	6,167	1,710	48,683	83,500
miR-382-5p	0,595	0,547	0,176	0,125	0,253	0,421	0,528	0,631	0,850	0,515	3,943	2,786	4,467	2,305
miR-485-5p	0,010	0,000	0,028	0,040	0,030	0,035	0,022	0,029	0,110	0,090	0,030	0,035	0,145	0,120
miR-506-3p	0,020	0,024	0,360	0,694	0,050	0,035	0,767	0,885	0,242	0,297	0,793	1,305	1,783	2,549
miR-508-3p	0,055	0,052	1,216	2,564	0,083	0,012	0,565	0,797	0,262	0,259	0,253	0,326	6,815	11,247
miR-509-3p	0,227	0,283	0,424	0,494	0,153	0,169	0,320	0,542	0,717	0,551	0,670	0,637	14,700	25,888
miR-510-5p	0,220	0,481	0,204	0,342	0,010	0,000	0,158	0,335	0,018	0,020	0,010	0,000	0,675	0,932
miR-514a-3p	0,032	0,053	0,368	0,801	0,010	0,000	0,430	0,600	0,050	0,098	0,010	0,000	3,503	7,250
miR-552-3p	0,025	0,037	0,024	0,031	0,010	0,000	0,038	0,063	0,050	0,052	0,010	0,000	0,112	0,096
miR-649	0,075	0,159	0,164	0,268	0,010	0,000	0,413	0,878	0,080	0,090	0,110	0,173	0,202	0,306
miR-650	0,085	0,184	0,388	0,845	0,010	0,000	0,222	0,391	0,077	0,103	0,380	0,332	0,537	1,160
miR-661	0,010	0,000	0,514	0,953	0,010	0,000	0,437	0,750	0,038	0,044	0,063	0,047	0,128	0,161
miR-873-5p	0,047	0,060	0,022	0,027	0,060	0,044	0,028	0,029	0,175	0,277	0,417	0,488	0,070	0,097
miR-9-3p	0,208	0,238	0,022	0,027	39,007	67,544	0,152	0,216	0,122	0,170	3,050	4,461	0,168	0,136
miR-92a-3p	39,217	22,356	77,540	25,437	54,633	51,962	106,517	80,730	106,917	67,092	118,000	59,573	234,483	174,523
miR-92b-3p	1,250	1,103	2,640	0,961	10,767	8,907	0,812	0,905	2,475	1,573	2,433	0,551	7,463	6,428
miR-96-5p	10,217	13,186	1,208	1,546	74,333	97,028	3,522	5,534	8,800	8,788	8,400	5,378	8,908	14,289
miR-99a-5p	31,383	24,957	144,040	162,070	109,367	178,101	133,350	197,812	93,767	73,884	228,000	192,299	263,800	168,560
miR-24-3p	143,850	73,771	321,600	115,355	316,500	453,512	528,417	941,975	273,517	231,553	771,667	461,461	374,167	305,332
miR-21-5p	520,952	1114,245	873,400	936,489	1240,003	1104,893	676,185	1144,425	1766,167	1844,084	3610,000	2022,325	8824,167	13593,237
miR-16-5p	81,033	42,960	338,000	103,634	127,867	131,203	830,600	1678,428	413,483	384,473	463,000	249,487	352,833	203,579
miR-320a	16,300	10,330	31,540	6,390	26,783	32,502	47,383	66,959	34,350	24,657	48,500	10,149	55,100	20,921
miR-224-5p	2,805	3,928	2,258	1,832	1,503	1,943	4,467	4,293	5,438	4,580	7,303	8,552	3,730	3,413
miR-423-5p	0,312	0,682	4,048	3,599	6,870	9,885	1,242	1,367	7,033	4,586	0,080	0,121	3,735	3,733
miR-25-3p	20,000	13,877	55,060	16,399	34,500	33,127	83,450	138,708	57,300	39,371	140,900	84,419	72,750	37,525
miR-331-3p	4,500	2,696	20,760	13,380	37,033	38,846	13,833	15,597	16,633	13,117	21,567	8,814	14,150	7,699
miR-103a-3p	34,160	9,669	108,920	58,780	122,733	83,948	95,750	97,950	121,017	102,493	170,000	85,913	169,933	100,222

	PAAD		PRAD		SKCM		TGSC		BLCA	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
let-7e-5p	11,836	16,083	59,482	88,652	11,900	19,110	1,890	2,014	19,903	25,025
let-7i-5p	179,880	133,715	296,060	380,399	57,333	92,059	22,575	11,417	167,733	198,689
miR-10a-5p	10,252	9,447	3,010	3,693	1,448	1,477	2,503	4,468	3,500	2,599
miR-10a-3p	0,230	0,232	0,210	0,243	0,113	0,115	0,040	0,036	0,068	0,085
miR-10b-5p	5,080	4,182	44,600	50,488	9,950	8,440	8,175	6,794	20,550	22,437
miR-106b-5p	17,620	14,281	73,640	83,655	21,029	28,535	53,725	44,402	13,375	7,899
miR-122-5p	0,118	0,205	0,193	0,139	0,033	0,039	0,010	0,000	0,070	0,120
miR-124-3p	0,312	0,255	0,660	0,822	0,216	0,352	13,708	12,214	0,355	0,500
miR-126-5p	0,914	0,732	2,458	3,190	4,504	5,584	3,425	1,223	3,305	2,597
miR-130a-3p	19,920	19,251	127,780	159,036	11,243	12,985	46,075	40,632	23,925	25,552
miR-135b-5p	43,034	46,792	13,428	26,046	0,901	0,720	36,950	30,319	5,133	4,706
miR-138-5p	0,102	0,126	0,025	0,030	10,333	17,467	0,263	0,426	2,053	2,557
miR-141-3p	66,940	30,187	515,660	589,402	17,717	13,824	14,175	10,634	28,975	18,526
miR-142-3p	75,680	91,051	267,900	327,492	111,350	107,051	149,500	110,386	70,100	89,552
miR-145-5p	802,160	780,128	3067,802	3102,922	58,229	40,593	12,960	13,669	2882,975	3506,579
miR-146a-5p	14,780	15,966	31,240	36,019	184,443	218,273	17,475	9,748	17,975	23,409
miR-148b-3p	7,020	4,633	31,100	40,956	3,059	3,623	2,873	2,115	4,050	2,929
miR-149-5p	1,626	1,882	65,340	83,555	1,690	0,758	2,393	1,969	10,925	7,638
miR-152-3p	24,080	19,781	89,920	115,041	8,787	7,514	2,520	1,548	30,700	36,969
miR-181a-5p	56,120	46,990	42,720	50,906	28,471	29,714	7,950	4,820	40,400	26,809
miR-181a-2-3p	1,764	1,816	1,540	2,038	1,336	1,730	0,038	0,032	0,620	0,865
miR-181b-5p	26,800	22,873	20,380	23,449	16,729	19,474	3,098	1,682	13,075	9,051
miR-182-5p	3,240	2,031	21,640	22,764	1,010	1,135	46,800	46,645	3,785	4,153
miR-183-5p	3,160	1,659	14,786	17,351	0,662	0,525	15,050	18,196	3,800	4,271
miR-187-3p	0,684	0,650	7,092	10,301	0,820	1,472	1,650	0,545	4,500	3,486
miR-187-5p	0,010	0,000	0,130	0,209	0,031	0,027	0,038	0,055	0,075	0,077
miR-192-5p	67,260	44,308	9,268	12,120	0,861	0,892	8,125	5,970	1,418	1,275
miR-193a-3p	4,276	4,671	4,886	6,829	2,103	2,133	3,550	2,810	1,753	1,411
miR-193b-3p	23,640	13,487	65,000	69,888	10,100	7,646	9,913	8,951	37,000	19,280

miR-194-5p	154,540	100,132	22,908	28,185	1,406	1,117	8,875	6,815	3,233	3,214
miR-194-3p	2,400	2,739	0,240	0,159	0,616	1,581	0,413	0,662	0,250	0,404
miR-196a-5p	3,606	6,302	1,186	1,691	3,497	2,648	2,285	3,822	2,425	2,816
miR-19b-3p	43,420	32,393	602,480	697,440	47,943	30,520	110,450	79,543	51,925	38,149
miR-200a-3p	16,080	11,570	53,482	99,813	2,107	2,292	0,040	0,060	3,725	3,084
miR-200a-5p	1,098	0,543	4,222	5,647	0,072	0,053	0,025	0,030	0,198	0,083
miR-200b-3p	145,500	87,638	1287,720	1999,225	7,200	4,864	0,228	0,160	38,675	36,247
miR-200b-5p	4,160	2,057	33,080	52,111	0,299	0,159	0,135	0,145	1,125	0,967
miR-200c-3p	141,560	65,314	870,200	1087,877	30,543	17,855	13,150	7,671	78,750	67,307
miR-204-5p	1,565	1,440	38,640	44,148	1,739	2,012	1,193	0,584	6,600	7,104
miR-205-5p	44,024	61,885	2167,500	2716,528	247,129	149,670	11,200	9,120	551,025	534,740
miR-210-3p	35,120	27,539	25,400	32,480	4,454	3,125	6,575	6,204	11,275	12,821
miR-211-5p	2,896	3,814	0,264	0,277	53,414	48,820	1,710	2,737	1,575	2,418
miR-214-3p	76,620	80,164	158,600	187,061	11,600	6,291	4,200	1,275	79,750	60,230
miR-215-5p	72,820	49,875	17,140	27,668	0,779	0,840	7,150	5,073	1,528	1,273
miR-27b-3p	29,184	60,328	96,960	108,082	181,944	267,681	10,525	5,236	118,725	101,305
miR-29a-3p	150,740	171,937	209,180	212,941	45,583	71,433	29,325	7,374	55,825	41,887
miR-29b-3p	109,440	152,890	302,920	342,605	41,350	57,242	17,275	10,196	33,175	17,041
miR-29c-3p	305,960	387,179	1080,060	1355,940	66,771	65,571	48,750	32,717	81,350	80,501
miR-30a-5p	13,680	9,126	63,000	69,470	14,571	19,345	3,500	1,971	15,225	19,711
miR-30a-3p	4,820	3,391	24,260	26,614	3,686	3,009	0,705	0,447	4,933	5,408
miR-30c-5p	138,700	120,447	670,480	788,959	63,100	46,044	65,700	38,833	139,400	161,846
miR-31-5p	59,520	83,971	46,128	68,390	4,130	3,943	8,063	7,952	110,650	79,300
miR-31-3p	3,830	4,391	2,318	3,234	0,507	0,623	1,158	1,164	7,775	3,925
miR-340-3p	1,200	0,648	2,406	2,522	0,737	0,937	0,873	0,400	0,825	0,171
miR-342-3p	66,680	100,166	152,500	174,675	28,743	22,310	21,000	5,957	38,775	42,901
miR-345-5p	11,020	8,952	39,760	42,952	3,754	5,355	8,828	6,840	8,550	6,030
miR-34a-5p	52,680	38,714	154,880	202,023	48,143	74,462	14,368	10,619	43,825	29,697
miR-34b-3p	0,146	0,190	0,240	0,241	0,085	0,111	0,083	0,055	0,168	0,092
miR-361-5p	21,920	19,092	95,880	112,637	14,250	15,394	18,600	14,540	29,300	28,356
miR-363-3p	3,920	2,389	167,060	214,702	3,476	4,138	1,323	0,469	1,775	0,737

miR-372-3p	0,446	0,601	0,212	0,325	0,126	0,217	549,275	359,020	0,325	0,409
miR-373-3p	1,236	1,891	0,072	0,045	0,043	0,033	543,700	419,397	0,240	0,208
miR-375	88,080	62,706	1041,180	1418,542	0,399	0,346	0,188	0,209	0,370	0,504
miR-382-5p	3,154	2,051	3,150	3,458	0,341	0,104	4,935	3,370	4,755	5,851
miR-485-5p	0,170	0,261	0,076	0,072	0,010	0,000	0,153	0,135	0,115	0,077
miR-506-3p	1,102	1,524	0,078	0,126	2,034	1,874	0,505	0,644	0,458	0,472
miR-508-3p	0,756	1,029	0,092	0,036	3,866	3,534	1,225	1,107	0,698	1,138
miR-509-3p	0,854	0,934	0,480	0,502	12,129	11,903	1,475	1,166	0,205	0,195
miR-510-5p	0,010	0,000	0,022	0,027	1,777	2,788	0,010	0,000	0,040	0,060
miR-514a-3p	0,022	0,027	0,010	0,000	2,520	4,004	0,243	0,212	0,050	0,080
miR-552-3p	0,098	0,197	0,258	0,365	0,010	0,000	0,220	0,163	0,280	0,226
miR-649	0,048	0,058	0,118	0,109	0,010	0,000	0,053	0,085	0,170	0,222
miR-650	0,092	0,058	0,010	0,000	0,029	0,033	0,053	0,057	0,210	0,139
miR-661	0,260	0,380	0,214	0,146	0,447	1,082	0,275	0,256	0,233	0,239
miR-873-5p	0,010	0,000	0,250	0,239	0,279	0,459	0,178	0,193	0,208	0,395
miR-9-3p	0,168	0,236	0,524	0,652	0,100	0,089	1,508	2,331	0,135	0,179
miR-92a-3p	78,700	44,375	300,880	297,385	70,586	73,176	131,725	100,459	97,300	49,401
miR-92b-3p	4,142	2,210	10,980	13,671	1,816	2,233	6,450	6,411	4,200	3,760
miR-96-5p	2,700	1,454	34,280	40,987	0,866	0,968	47,375	37,595	1,995	2,101
miR-99a-5p	134,480	175,809	2473,000	3066,442	28,114	20,322	24,025	15,061	461,650	704,957
miR-24-3p	485,200	386,420	1492,000	1984,229	394,786	471,495	40,700	10,249	448,500	371,003
miR-21-5p	3020,000	2256,029	3566,200	6071,825	1375,067	1960,302	525,500	318,459	706,000	491,651
miR-16-5p	214,660	162,347	909,020	1053,710	322,429	374,779	186,000	54,006	257,050	284,162
miR-320a	65,460	29,817	166,720	235,768	30,871	24,541	44,050	42,323	94,025	81,696
miR-224-5p	4,124	3,857	5,086	8,398	0,776	1,088	0,878	0,268	6,055	7,868
miR-423-5p	4,042	3,443	7,512	10,917	3,671	5,967	2,725	1,493	3,328	2,550
miR-25-3p	52,020	35,748	147,280	161,334	62,829	78,004	159,450	86,853	40,267	33,913
miR-331-3p	23,360	17,345	51,800	64,580	11,444	16,755	6,105	4,395	18,925	13,732
miR-103a-3p	156,400	125,781	472,220	605,080	100,700	134,150	56,725	44,737	110,350	96,020

Supplementary Table 2. Error rates of the two models for each tumor class.

Tumor class	PAMR	LASSO
BLCA	0,75	0,70
OV	0,74	0,46
PAAD	0,69	0,72
TNBC	0,67	0,53
CHOL	0,50	0,38
HNSC	0,49	0,47
LIHC	0,37	0,48
KICA	0,25	0,31
TGSC	0,25	0,25
GI-NET	0,20	0,28
UCEC	0,18	0,00
STAD-CRC	0,16	0,16
LUSC	0,16	0,48
SKCM	0,14	0,06
LUAD	0,08	0,04
LBC	0,03	0,36
PRAD	0,02	0,00