16

Università degli Studi di Ferrara

DOCTORAL COURSE IN

Biomedical Sciences and Biotechnology

CYCLE XXXIV

DIRECTOR Prof. Paolo Pinton

Novel molecular and genetic approaches to support the clinical management of Cancers of Unknown Primary (CUPs)

Scientific/Disciplinary Sector (SDS): MED/04

Candidate
Dott.ssa Noemi Laprovitera

_____
(signature)

Supervisor
Prof.ssa Silvia Sabbioni

_____
(signature)

Years 2021/2022

INDEX

SOMMARIO

Le metastasi sono responsabili della maggior parte delle morti cancro-relate. L'iter diagnostico standard nella maggior parte dei casi è in grado di riconoscere il sito di origine dei tumori metastatici, ma in alcuni casi questo rimane ignoto, portando a una diagnosi di tumore di origine primaria sconosciuta (*cancer of unknown primary site*, CUP). I CUP rappresentano un gruppo eterogeneo di tumori metastatici rari (circa il 3-5% di tutte le diagnosi di cancro) caratterizzati da un comportamento clinico aggressivo e da una prognosi infausta.

Poiché le linee guida sono tumore-specifiche, la mancanza di un sito primario noto impedisce ai pazienti CUP l'accesso a diversi tipi di terapia e a trial clinici, e questo ha un forte impatto sulla loro aspettativa di vita. Inoltre, come recentemente osservato, i CUP presentano alterazioni che potrebbero renderli eligibili a terapie mirate o immunoterapia. Tuttavia, l'analisi mutazionale nei pazienti CUP è spesso ostacolata dalla scarsità del materiale bioptico tumorale a seguito dell'approfondito iter diagnostico e dalla mancanza di disponibilità commerciale di pannelli di geni dedicati ai CUP.

In questo studio, abbiamo analizzato il profilo di 89 microRNA (miRNA) per dedurre il sito di origine dei tumori metastatici, inclusi 53 campioni CUP. L'espressione dei miRNA è stata misurata mediante tecnologia Droplet Digital PCR (ddPCR) in 159 campioni fissati in formalina ed inclusi in paraffina (FFPE), che includevano tumori primari appartenenti a 17 classi tumorali (set di riferimento), metastasi di origine nota e CUP. Utilizzando due diversi modelli statistici per la predizione del sito di origine, in particolare PAMR (*The Nearest Shrunken Centroids Approach of Prediction Analysis of Microarrays*) e LASSO (*Least Absolute Shrinkage and Selection Operator*), abbiamo ottenuto con successo una previsione molecolare per tutti i campioni FFPE testati. Le sedi primarie più frequentemente predette sono state tratto gastrointestinale, pancreas, seno, polmone e dotto biliare. Il test è stato applicato anche a metastasi multiple sincrone e metacrone derivanti dallo stesso paziente CUP: le previsioni hanno mostrato un forte accordo, convalidando intrinsecamente il nostro test. Le predizioni molecolari sono state poi confrontate con le prime ipotesi cliniche e/o patologiche del sito d'origine, ove disponibili. La nostra analisi ha inoltre individuato 13 miRNA come potenziali biomarcatori prognostici dei pazienti CUP.

Nella seconda parte dello studio, abbiamo analizzato il potenziale della biopsia liquida per la caratterizzazione genetica di due pazienti CUP prospettici, utilizzando

sia cellule tumorali circolanti (CTC) che DNA libero circolante (ccfDNA), per identificare alterazioni genetiche che li rendano eligibili a terapie mirate. Le CTC, isolate dai campioni di sangue dei pazienti CUP utilizzando i sistemi CELLSEARCH®/DEPArray™ NxT e Parsortix, sono state caratterizzate dal punto di vista immunofenotipico e utilizzate per l'analisi genomica a singola cellula con il kit Ampli1™. Inoltre, il ccfDNA è stato purificato da campioni longitudinali di plasma per l'analisi genetica utilizzando un pannello personalizzato di 92 geni dedicato ai CUP mediante tecnologia SureSelect di *Target Enrichment*. Parallelamente, il DNA del tumore è stato analizzato con tre diversi pannelli NGS: FoundationOne CDx (F1CDx), DEPArray OncoSeek e il pannello personalizzato SureSelect. Quando il gene o la regione erano coperti dal pannello, abbiamo ottenuto un'elevata concordanza nelle mutazioni rilevate, fatta eccezione per una inserzione nel gene APC che non è stata osservato da F1CDx. In un paziente abbiamo identificato l'amplificazione dei geni *FGFR2* e *CCNE1* nelle singole CTC, nel DNA tumorale e nei ccfDNA. Inoltre, una mutazione puntiforme nel gene ARID1A (p.R1276*) è stata riscontrata nel DNA tumorale e nei ccfDNA. Le alterazioni sono state validate mediante ddPCR in tutti i campioni di ccfDNA raccolti longitudinalmente durante il decorso della malattia. L'analisi delle CTC di un secondo paziente ha mostrato pattern ricorrenti di amplificazione nei geni ASPM e SEPT9 e delezione in FANCC. Il pannello personalizzato SureSelect ha individuato 16 alterazioni nel ccfDNA, tra cui una delezione (I1485Rfs*19) e una mutazione somatica (p. A1487V) nel gene ARID1A e una mutazione puntiforme nel gene *FGFR2* (p.G384R).

Il nostro studio ha dimostrato che il profilo di espressione dei miRNA nei campioni CUP può essere eseguito su richiesta, in associazione con il workup diagnostico standard, per offrire preziose indicazioni sul possibile sito primario. Tale strumento molecolare, combinato con approcci di biopsia liquida per l'analisi mutazionale, potrebbe supportare le decisioni terapeutiche, offrire nuove opzioni di trattamento per i pazienti CUP e potenzialmente migliorare la loro aspettativa di vita.

Abstract

Metastasis is responsible for the majority of cancer-related deaths. Standard diagnostic workup usually is able to recognize the tissue-of-origin (TOO) of metastatic tumors but in some cases, the TOO remains undetermined, leading to a diagnosis of Cancer of Unknown Primary origin (CUP). CUPs comprise a heterogeneous group of rare metastatic tumors (approximately 3-5% of all cancer diagnoses) characterized by an aggressive clinical behavior and dismal prognosis. Since treatment guidelines are tumor-gnostic, the lack of a defined TOO precludes CUP patients from accessing to therapeutic options or clinical trial, significantly impacting their life's expectancy. Moreover, as recently reported, CUPs presents potentially actionable alterations for which targeted therapies or immunotherapy could be proposed. However, CUP genetic testing is hampered by the paucity of tumor biopsy material after the extensive diagnostic workup and the lack of CUP-dedicated commercial gene panels for target sequencing.

In this study, we applied a pre-determined set of 89 microRNAs to predict the TOO of metastatic cancers, including 53 CUP samples. The miRNA expression was assessed with Droplet Digital PCR (ddPCR) in 159 formalin-fixed paraffin-embedded (FFPE) tissues samples, including primary tumors from 17 tumor classes (reference set), metastases of known origin (test set) and CUPs. Using two different statistical models for class prediction, specifically the nearest shrunken centroids approach of Prediction Analysis of Microarrays (PAMR) and the Least Absolute Shrinkage and Selection Operator (LASSO), we successfully obtained a molecular prediction for all the FFPE samples tested. The most frequently predicted TOO were gastrointestinal, pancreas, breast, lung and bile duct. The assay was applied also to synchronous and metachronous multiple metastases derived from the same CUP patient: the predictions showed a strong agreement, intrinsically validating our assay. Final molecular predictions were also compared with the first clinical and/or pathological hypotheses of a primary site, when available. Moreover, our analysis reported also 13 miRNAs as potential prognostic biomarkers of CUP patients.

In the second part of the study, we analyzed the potential utility of liquid biopsy for the genetic characterization of two prospective CUP patients, using both circulating tumor cells (CTCs) and circulating cell-free DNA (ccfDNA), to identify druggable alterations. CTCs, isolated from CUPs' blood samples using the CELLSEARCH®/ DEPArray™ NxT and Parsortix systems, were immunophenotypically characterized and used for single-cell genomic characterization with Ampli1™ kits. In addition,

ccfDNA was purified from longitudinal plasma samples for genetic testing using a CUP-dedicated, 92-gene custom panel and SureSelect Target Enrichment technology. Parallely, the bulk tumor DNA was analyzed with three different NGS panels: FoundationOne CDx assay (F1CDx), the DEPArray OncoSeek Panel, and the SureSelect custom panel. When the gene or region was covered by the panel, we obtained high concordance in the mutations detected, except for an insertion in APC gene which was not observed by F1CDx. Amplifications in FGFR2 and CCNE1 genes were identified in single CTCs, tumor tissue and ccfDNAs in one patient. A point mutation in ARID1A gene (p.R1276*) was detected in the bulk tumor DNA and ccfDNAs. The alterations were validated by ddPCR in all ccfDNA samples collected during tumor evolution. CTCs analysis from a second patient exhibited a recurrent pattern of amplifications in ASPM and SEPT9 genes and deletion of FANCC. The SureSelect custom panel reported 16 somatic alterations in ccfDNA, including a deletion (I1485Rfs*19) and a somatic mutation (p. A1487V) in ARID1A gene and a point mutation in FGFR2 gene (p.G384R).

Our study demonstrated that miRNA expression profiling in CUP samples could be can be performed on demand, in association with standard diagnostic workup, to offer valuable indications about the possible primary site. Such molecular tool, combined with liquid biopsy approaches to identify druggable genetic alterations, could assist treatment decisions, offer novel therapeutic options for CUP patients, and potentially prolong their life expectancy.

Keywords: microRNAs, cancer of unknown primary, droplet digital PCR, liquid biopsy, circulating tumor cells, circulating tumor DNA

List of abbreviations

CUP, cancer of unknown primary site; TOO, tissue of origin; FFPE, formalin-fixed, paraffin-embedded; CTC, circulating tumor cells; ccfDNA, circulating cell-free DNA; ctDNA, circulating tumor DNA; IHC, immunohistochemistry; GEP, gene expression profiles; miRNA, microRNA; GEP, gene expression profile; ddPCR, droplet digital PCR; F1CDx, FoundationOne CDx assay; BLCA, transitional cell carcinoma of bladder; CHOL, cholangiocarcinoma; CRC, colorectal adenocarcinoma; GI-NET, gastrointestinal neuroendocrine carcinoma, HNSC, head and neck squamous cell carcinoma; KICA, kidney renal clear and renal papillary cell carcinoma; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LBC, luminal non-special type and lobular breast carcinoma; LIHC, hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OV, ovarian serous carcinoma; PAAD, pancreas exocrine adenocarcinoma; ; PRAD, prostate adenocarcinoma; SKCM, melanoma of skin; STAD, gastric adenocarcinoma; TGSC, germ cell seminomatous carcinoma; TNBC, triple negative breast cancer; UCEC, endometrial adenocarcinoma; ND, not defined; HE, hematoxylin-eosin; PAMR, Prediction Analysis of Microarrays for R; LASSO, least absolute shrinkage and selection operator; NSC, Nearest shrunken centroids; BRCA, Breast invasive carcinoma; TCGA, The Cancer Genome Atlas; OS, Overall survival; PPR, positive prediction rate; ROC, receiver operating characteristic; HR, hazard ratio; HPV, human papilloma virus.

Introduction

The emergence of metastasis represents the primary cause of death for more than 90% of cancer patients and affects their response to therapy and life expectancy [1,2]. Cancer metastasis consists in the spread of cancer cells to tissues and organs distant from the primary tumor site to originate secondary tumors. “Activating invasion and metastasis” represents an important hallmark of cancer and is a key feature of cancer malignancy [3]. However, the development of metastases involves a number of sequential events to be successful, defined as “metastatic cascade” (**Figure 1**) [4]. It requires tumor cells to leave their primary site, circulate in the bloodstream, endure pressure in blood vessels, adjust to new cellular surroundings in a secondary site and escape immune surveillance.

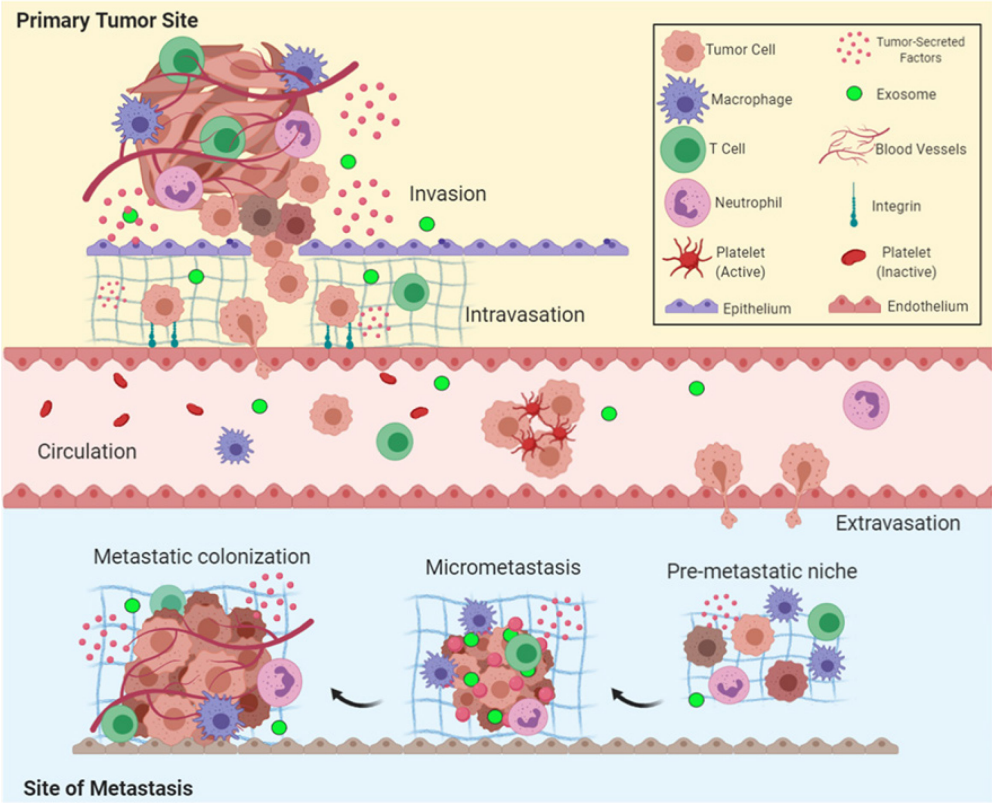

Figure 1. Overview of the metastatic cascade: the main steps of metastasis include invasion, intravasation, circulation, extravasation, and colonization [4].

This process is caused by the progressive acquirement of genetic alterations in tumor cells, which eventually become anchorage independent and self-sufficient for cell growth. The worst-case scenario of a metastatic disease is represented by cancer of unknown primary site (CUP), a heterogeneous disease diagnosed -by definition- at an advanced stage.

CUPs, also known as “occult primary tumors” or “metastases of unknown origin”, comprise histologically confirmed metastatic tumors, whose primary site is not determined after extensive diagnostic clinical and pathological investigation [5]. The first mention in literature of cancer with unknown primary dates back to 1946 [6]. However, despite a substantial improvement of the diagnostic tools and the extensive workup to seek the primary site, CUP remain to date an important clinical problem, specifically when considering that current therapeutic guidelines are primary site oriented, and its biology is still an enigma [7].

At initial presentation, 50% of CUP patients have multiple metastases that involve two or more organs [8]; percentages of most common metastatic sites observed at diagnosis in three historic CUP cohorts ([9-11]) were reported in **Figure 2A** along with those observed in 44 CUP patients enrolled in a currently ongoing clinical trial (NCT02607202). Although the numerical differences, lymph node, bone, liver and lung metastases remain the most common metastatic sites.

Despite their heterogeneity, a set of distinct recurring features characterizes CUPs, comprising early dissemination, aggressiveness, an atypical metastatic organotropism (**Figure 2B**), poor response to therapy and, consequently, short life expectancy [12,13].

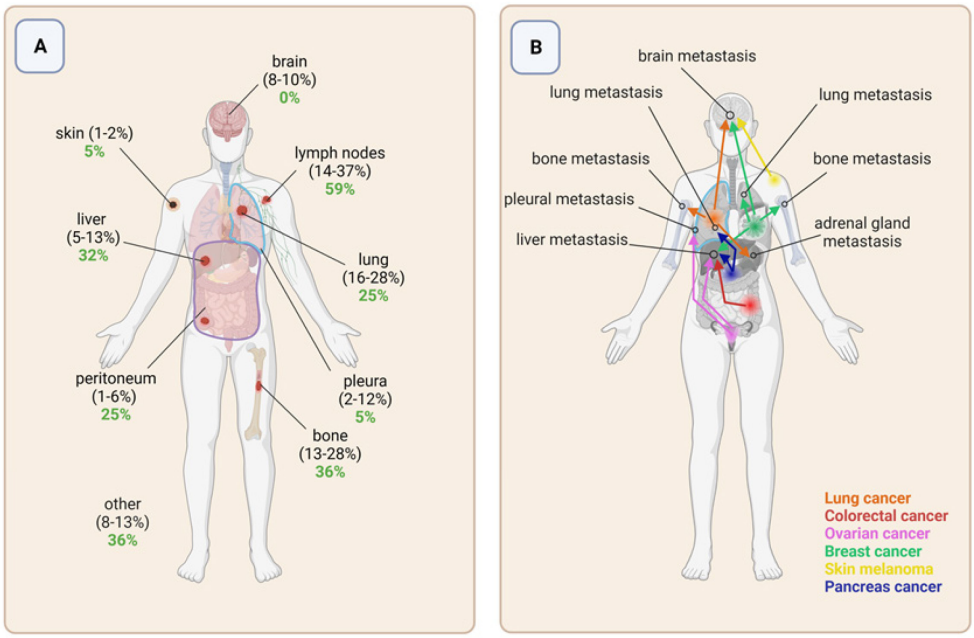

Figure 2. Metastatic sites and organotropic metastatic patters of CUP. A) Schematic representation of most common metastases at diagnosis of CUP patients. The figure reported a comparison of data from three published historic cohorts (in brackets, [9-11]) with that of an ongoing clinical trial (in green, [14]). **B)** Representation of most common metastatic patterns in CUP patients [13]. Created with BioRender.com, adapted from [14] and [13].

Epidemiology

CUP was historically reported as 3-5% of novel worldwide cancer diagnoses [5]. However, recently CUP incidence rate has decreased to 1–2% [15], even though still ranking among the ten most frequent and deadly tumors in the world [16].

Six population-based studies on CUP incidence (reviewed in [15]) reported an initial increasing trend in incidence since CUP concept was introduced, followed by a decline up to nowadays (**Figure 3**) possibly as a consequence of the availability of novel more sophisticated diagnostic techniques aimed to trace the primary tumor.

Of note, all five European Union countries [17-21] and Australia [22], present the highest peak between 1990-2000 [15]. CUP incidence is varying across different countries. Differences in incidence trends among these population-based studies may be explained by different coding rules and methodology in cancer registries [21].

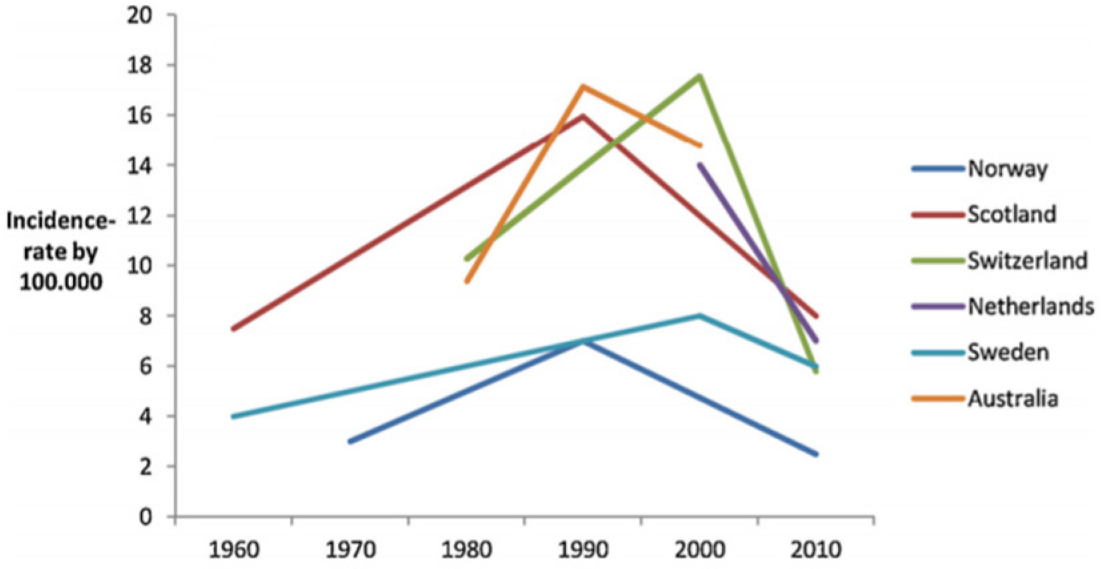

Figure 3. Variations in incidence trends of six different countries. This graph shows approximate CUP incidence trends of six countries. It is recurrent the presence of the highest peak between 1990-2000 and a decreasing trend in the last decade [15].

On the contrary, in United States, the highest incidence peak occurred in 1980s and its incidence-rate has been decreasing of 3.6% each year in the last decades [23]. However, cost issues can explain underreporting of CUP in the US: the costs are better covered for specific cancers in comparison with CUPs [23]. In addition, the decreasing number of CUP diagnoses could be explained not only by a better identification of small primary lesions, but also by a decreased incidence of some primary cancers due to improved lifestyles or screening programs.

Differences in CUP incidence rate are also biased by socioeconomic disparities between countries and by the unequal access to diagnostic workup and treatment [23,24]. Therefore, since it is a real challenge to compare studies from different decades and countries, further epidemiological studies are needed to shed light on this enigmatic disease.

CUP occurs mainly in adult patients with a median age of 60 years and a slightly higher prevalence in male population [25]. Even if rarely, it may also occur in adolescents and young adults (15-39 years) representing a real clinical challenge and a condition that is still poorly understood and investigated [26,27].

Despite the hypothesis of a "familial CUP" should be further explored, few familial clustering studies reported that CUP patients' first-degree relatives have a major risk of CUP and other tumors (lung, kidney, liver, ovarian, colorectal, breast and melanoma) compared to general population [28]. In addition, other known risk factors associated with CUP are smoking habit [29] and human papillomavirus (HPV) infection for squamous cell CUPs [26,30].

The pathogenesis of CUP

The mechanism underlying CUP origin to date is still poorly understood. To explain the reason why the primary tumor remains elusive during the diagnostic work-up, the scientific community have proposed several hypotheses and mechanisms that could unveil CUP biology [31].

According to the two prevailing hypotheses, CUPs could originate from small undetectable, dormant, or later regressed primary lesions or represent early disseminating, aggressive metastatic entities with no existing primary site [5,32].

As previously mentioned, according to current knowledge, the development of metastases requires sequential events known as "metastatic cascade". However, in the context of CUP, metastases might occur before local tumor growth.

In this case (**Figure 4**), cells could spread at an early stage to distant metastatic sites and modify their microenvironment, forming a metastasis before the primary tumor becomes detectable or even before the pre-malignant lesion acquires a fully malignant phenotype [33,34]. This theory is supported by the "parallel progression model": tumor cells from the primary tumor and from metastatic sites have independent clonal evolution, therefore metastatic cells show no genetic similarity with the primary lesion and with other metastases derived from the same patient [35,36].

To support this theory, a recent study demonstrated a higher chromosomal instability (CIN) in CUP patients if compared with metastases of known origin or primary tumors. CIN, combined with parallel metastatic behavior, could be linked to early dissemination, poor outcome and drug resistance that characterize CUPs [37]. In addition, to describe CUP aggressiveness, it was also hypothesized that in these patients the step of “pre-metastatic niche formation”, a process in which the primary tumor prepares the environment for metastatic sites by secreting specific factors, is bypassed. Indeed, CUP metastatic cells seem to be independent from the primary site and able to start the metastatic process without its contribution [38,39].

Another theory sees metastases formation occurring through the normal tumor progression, in which the metastatic cascade is the final stage of tumorigenesis. In this case, the tumor microenvironment might play a pivotal role selectively favoring the outgrowth of tumor cells at the metastatic site and inhibiting the primary site (**Figure 4**) [40,41]. The evasiveness of the primary tumor, in this case, is attributed to its small size, its disappearance by regression or its state of dormancy. In fact, primary tumors usually identified by autopsy are mainly of small size [31].

This hypothesis, defined as the “linear progression model” is supported by evidence of a clonal relationship between cells of the primary and metastatic sites [42]. Due to the molecular similarity observed between primary and metastatic tumors documented in different studies, it was hypothesized that these patients could have benefited from site-specific therapy [31]. To support this theory, a recently published case report described the results of a comprehensive genetic and transcriptomic analysis of multiple metastases derived from the same CUP patient [43]; this study reported a singular homogeneity between metastases that suggested the presence of a common ancestor that linearly and rapidly accumulated mutations, and disseminated a progeny at subsequent stages of its evolution.

One of the hypothesized mechanisms that allows tumor cells to migrate and generate metastases is the epithelial-to-mesenchymal transition (EMT). This complex process consists in the loss of tumor cells of the epithelial phenotype (partial loss of E-cadherin expression, upregulation of N-cadherin/vimentin along with SNAIL) and the acquisition of mesenchymal features (upregulation of metalloproteinases) that allow them to enter the bloodstream or lymphatic system. However, by cell staining the EMT phenotype was uncommonly seen in tumors from CUP patients (7.3-16%), probably due to the transient nature of this phenomenon [44,45].

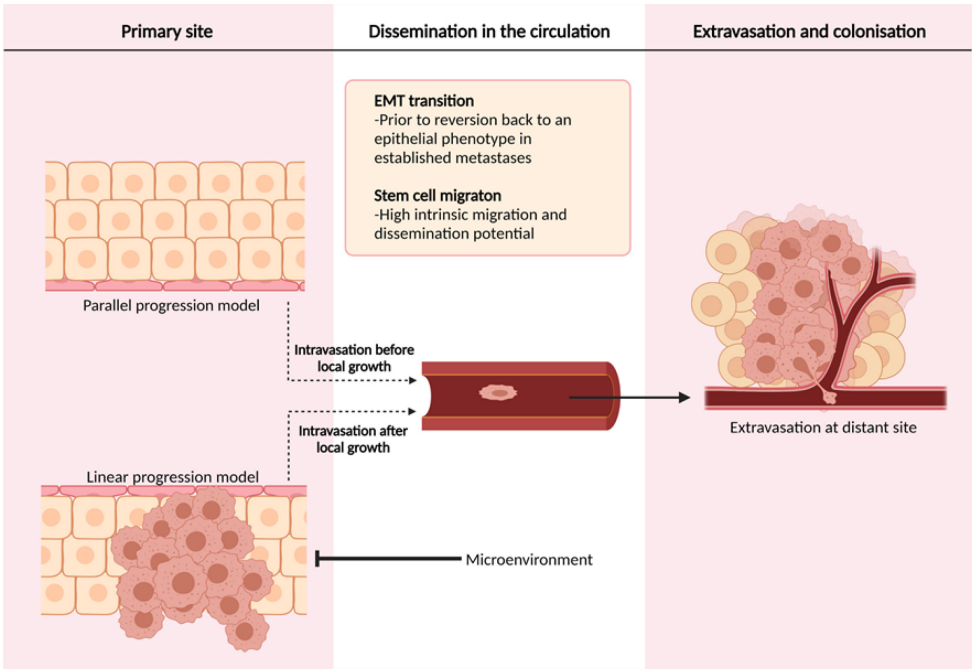

Figure 4. Potential mechanisms for the carcinogenesis of CUP. This figure (created with BioRender.com, adapted from [36]) represents linear and parallel tumor evolution models to generate metastasis.

Another recent hypothesis sees cancer stem cells, which are characterized by a high intrinsic migration/dissemination potential, to form CUP metastases [34,46]. The stem cell phenotype, defined by the immunohistochemical expression of CD133 and OCT4, might play a key role in the rapid dissemination of CUP.

The etiology of CUP is still a debated topic and should be further explored to clarify the underlying mechanisms. However, to date, no specific point mutations in oncogenes or suppressor genes have been found associated with CUP etiology [47].

CUP diagnosis

When dealing with a histologically confirmed metastasis with no evidence of a detectable primary site, clinical practice guidelines suggest a thorough diagnostic workup prior to establish a diagnosis of CUP [5,48].

Standard diagnostic workup consists of baseline and additional investigation and pathological evaluations.

Baseline investigation includes the evaluation of patient medical history, extensive physical examination, blood/biochemical analyses and radiological tests (chest radiography and computed tomography -CT- for thorax, abdomen and pelvis).

If the first series of diagnostic tests fail to identify the primary tumor, additional specific tests can be performed according to the localization of the metastases, patient's signs, symptoms and gender. These tests include magnetic resonance imaging (MRI), endoscopy, positron emission tomography (PET), mammography or the assessment of specific serum tumor markers.

Moreover, the pathological characterization of metastases by immunohistochemical (IHC) testing is the final crucial step to establish the tissue of origin, based on the assumption that there is a similarity in the expression profiles of primary and metastatic cancers.

If after these attempts the primary tumor is not identified, a diagnosis of CUP is confirmed. However, this condition is reversible since in some cases the primary tumor is unveiled later during the course of the disease. In addition, many circumstances can lead to a misdiagnosis of CUP such as a rushed premature diagnosis, the inexperience of the pathologist or the limited availability of tissue that hampers the number of performable diagnostic tests.

In addition, post-mortem investigations on CUP patients revealed a primary tumor in about 75% of cases and highlighted the prevalent epithelial origin of CUPs [49]. According to twelve cohort autopsy studies, five published between 1977-1980 [50-54] and seven between 1981-2005 [11,55-60], the most common primary sites identified in CUP patients were lung, pancreas, hepatobiliary tract, kidney, colon, genital organs and stomach [49].

Histological and immunohistochemical evaluation

Once the biopsy is obtained and the malignancy is established, the pathologist uses a stepwise approach to identify the broad tumor type, the tumor subtype and, whenever possible, the tissue of origin.

The differences between distinct tissue types depend on their gene expression profiles. Some active genes, involved in basic cellular functions, are ubiquitous, while others are expressed by one or few tissue types and impact their function and differentiation [61]. As mentioned before, the expression of a subset of genes is retained in metastatic cells and can be used to determine the primary tumor. Gene expression profiles in poorly or undifferentiated cancers can be preserved less efficiently than well-differentiated cancers [62]. However, despite the expression of a particularly informative gene might be lower in metastases compared to primary tumors, it could provide a decisive clue to confirm the suspected tissue of origin; for

this reason, immunohistochemical (IHC) testing covers a key role in pathological evaluation of metastases.

The first IHC staining is meant to define the tumor lineage and exclude highly treatable tumor types such as lymphoma, melanoma, germ-cell tumors or sarcoma (**Table 1**). If an epithelial lineage is confirmed, which is the case of the majority of CUPs, the pathologist evaluates the expression of two keratins (also known as cytokeratins), K7 and K20, which can narrow down the subsets of carcinoma, prior to assess other more specific markers [63,64].

These keratins are intermediate filaments that derive from 54 functional genes in the human genome and are expressed by epithelial cells [65]. While K7 expression is shared across different simple epithelia, K20 is restricted to specific epithelia. However, beyond this general rule, some carcinomas are known for their lack of immunophenotypes: i.e. pancreatic carcinoma generally is K7+/K20+, but it can also be K7+/K20-; gastric adenocarcinoma can show all K7/K20 phenotypes; cholangiocarcinoma can overlap with pancreatic carcinoma [66-68]. In addition, among carcinomas, IHC will distinguish three histopathological subtypes: adenocarcinoma (~50% of cases), poorly differentiated carcinoma (~30% of cases) and squamous cell carcinoma (~15% of cases) [69].

Once established one of the four K7/K20 expression patterns ($K7^+/K20^+$; $K7^+/K20^-$; $K7^-/K20^+$; $K7^-/K20^-$), IHC testing can proceed with disease-specific markers.

Table 1. Initial diagnostic immunohistochemistry flowchart for patients with cancer of unknown primary [70].

Step 1: define the lineage	**Step 2: define the subtype**	
Positive pancytokeratin (AE1/AE3, Cam5.2, OSCAR, etc.)	K7+/K20– ➔	Breast carcinoma
		Cholangiocarcinoma
		Endometrial adenocarcinoma
		Endocervical adenocarcinoma
		Gastric adenocarcinoma
➔ *Carcinoma*		Lung adenocarcinoma
		Mesothelioma
		Ovarian (serous) carcinoma
		Pancreatic adenocarcinoma
		Renal (papillary)
		Salivary gland tumors
		Small cell lung carcinoma
		Thyroid carcinoma
		Urothelial carcinoma (subset)
	K7+/K20+ ➔	Bladder adenocarcinoma
		Cholangiocarcinoma
		Gastric adenocarcinoma

		Ovarian mucinous carcinoma
		Pancreatic adenocarcinoma
		Urothelial carcinoma
	K7–/K20+ ➔	Colorectal adenocarcinoma
		Gastric adenocarcinoma
		Merkel cell carcinoma
	K7–/K20– ➔	Adrenocortical carcinoma
		Gastric adenocarcinoma
		Hepatocellular carcinoma
		Mesothelioma
		Non-seminoma germ cell tumors
		Prostate adenocarcinoma
		Renal (clear cell types)
		Small cell lung carcinoma
Positive HMB-45/Melanin-A/S100/SOX10 ➔ *Melanoma*	More testing may be needed to determine subtype	
Positive CD45/CD20/CD3 ➔ *Lymphoma*	More testing may be needed to determine subtype	
Other ➔ *Sarcoma*	More testing may be needed to determine subtype	

A hand-picked combination of cancer-specific stains is tested to confirm the suspected tissue of origin [5,69] as described in **Table 2**. The number of available IHC stains is always expanding as more specific antigens are identified. Despite the existence of specific practice guidelines for CUP IHC testing [48], pathologist's intuition and experience in markers selection is fundamental in the optimal management of the available tissue material. In addition, this process lacks of objective parameters, is time-consuming, complex and can lead to ambiguous classifications, especially when dealing with poorly differentiated tumors.

Table 2. Cancer-specific immunohistochemistry stains [70-72]

IHC marker	Bladder	Breast	Colorectal	Gastric	Lung	Mullerian	Pancreas/biliary	Prostate	Germ cell	Hepatocellular	Neuroendocrine	Renal (clear cell)	Renal (papillary)	Small cell lung	Squamous cell	Thyroid	Undifferentiated
	Adenocarcinoma								Carcinoma								
AE1/AE3, Cam 5.2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+/–
Arginase/HepPar1										+							
CD10/RCC												+	+				
CDX2		–	+	+	–		+/–	–	–		–					–	–
CK5/CK6															+		
CK7	+	+	–	+/–	+	+/–	+	–		–	+/–	–	+	+/–	+/–	+	–
CK20	+	–	+	+/–	–	–	–	–		–	–	–	–	–	–	–	
Chromogranin, NSE, synaptophysin		–	–		–			–	–		+			+		+/–	–

ER		+/–	–		–	+		–	–		–					–	–
GATA3	+	+													+		
GCDFP-15, mammaglobin		+															
Napsin A					+	+/–		–	–		–					–	–
NKX3-1	–	+/–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
OCT4, PLAP, SALL4									+								
p40, p63	+														+		
PAX8						+						+	+			+	
PAP, PSA	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
SATB2			+														
Thyroglobulin		–	–		–			–	–		+/–					+	–
TTF1		–	–		+			–	–	+	+/–					+	–
Villin			+		+/–		+/–										

CUP classification and treatment

International guidelines for tumor treatment are basically tumor-oriented. Therefore, with no indication of a primary site, CUP patients cannot be treated with cancer-specific therapy, so their management is challenging and basically depends on their clinicopathological characteristics. As a consequence, CUPs are usually treated with an empiric blind approach using platinum-based chemotherapy regimens that are poorly effective, resulting in a short life-expectancy for these patients (average overall survival 4-9 months, 20% survive more than 1 year) that have not improved in the last decades.

According to treatment response these patients can be classified into two prognostic subgroups: favorable (15-20%) and unfavorable CUPs (80-85%) [5,73]. Favorable CUPs share some features with a specific known tumor type, such as its metastatic pattern, chemosensitivity and prognosis.

These patients present a potentially curable disease if treated in line with similar tumor types and have a life expectancy of 15-20 months and can achieve a long-term disease control in 30-60% of cases [74].

CUP patients with cervical lymph node metastases from squamous cell carcinoma or peritoneal carcinomatosis from papillary serous carcinoma belong to this subgroup [74].

On the other hand, the majority of CUPs belong to the unfavorable group and have a short life-expectancy (6-10 months) [75]. According to a prognostic model developed by the French CUP Group (GEFCAPI), unfavorable CUPs can be further classified based on two parameters: the performance status and the pre-treatment serum lactate dehydrogenase levels (LDH) [76].

CUP patients characterized by a good performance status (0-1) and normal LDH levels, are responsive to platinum-based doublet chemotherapy regimen (in combination with gemcitabine or taxane) and achieve a median survival of 12 months [5,77].

On the contrary, patients with a performance status ≥ 2 and high LDH levels have a dismal prognosis (median survival of 4 months) and their treatment manly consists in palliative care, symptom control or low-toxicity chemotherapy regimens [48,73].

Of note, no chemotherapy agent (across platinum, taxane or new-generation compounds such as gemcitabine, vinca alkaloids or irinotecan) was found to effectively prolong CUP patient survival in monotherapy or in combination [12].

Nevertheless, the NCCN guidelines reported up to 11 chemotherapy regimens indicated for adenocarcinoma and 9 for squamous histology [48], whose application remains empirical since is based on single-arm phase II clinical trials [78-80] or small trials [81-83].

Molecular prediction of the primary site

Due to the tissue-gnostic setting of current treatment guidelines and the assumption that primary-oriented treatments would significantly affect CUP prognosis, several research groups in the past decade explored different approaches to obtain an indication of the primary site. These approaches based on mRNA, microRNA or DNA methylation analysis collected evidence that supported their translational potential in the clinical management of CUP patients. They are based on the principle that, independently by the technology used, metastatic tumors retain tumor specific molecular features during metastatic process.

Although based on different multi-feature algorithms, these molecular predictors are pretty similar in setting: at first a cohort of primary tumors (reference or training set) is analyzed to create a cancer-specific signature database that is employed as a tissue-of-origin classifier; then, based on the common features with the tumors of known primary site of the training set, a tissue-of-origin is assigned to each metastatic tumor with uncertain/unknown origin.

The accuracy of each molecular assay can be directly estimated by analyzing the same primary tumors used to create the classifier, independent datasets of primaries or metastatic tumors of known origin (cross validation or the employment of a test set). The reported prediction accuracy of these molecular profiling tools ranges between 80-95%, thus showing a great potential if applied in the diagnostic workup of CUP patients [47].

However, some limitation can be recognized in these assays: they are based on a low number of cancer-types or samples for each class in the reference set, employ large-scale analyses that are expensive (microarray-based studies), have a low number of analyzed targets, do not always succeed in the primary site prediction due to technical issue caused by low-quality RNA (GEP-based studies), can generate confound results due to contamination of tumor biopsy with surroundings tissues.

Moreover, prediction consistency of the site-of-origin in CUP patients, in most cases cannot be verified, given that the identification of the real primary site during the

course of the disease or at post-mortem autopsy is seldom available. In most cases the prediction is evaluated using available surrogate measures (comparison with IHC findings, clinical presentation, response to therapy).

Since to date, randomized, prospective trials did not reported robust evidences of a clinical benefit for CUP patients from the use of molecularly-directed therapies, current international and national guidelines do not recommend a routine use of tissue-of-origin predictors in CUP clinical management but a case-by-case evaluation [47,48].

Predictive assays based on gene expression profiling

With the goal to overcome the limits of current cancer classification, establish a more accurate molecular diagnosis and potentially improve the choice of CUP therapeutic options, several researchers focused in the development of molecular assays able to classify human cancers [84-86]. This strategy is based on the observation that metastatic tumor cells maintain some molecular characteristics of the tissue of origin, despite going through de-differentiation and epithelial-mesenchymal transition programs. So, this tissue-specific molecular signature can be used to infer CUPs' sites of origin.

Several gene-expression profile (GEP) -based molecular tests were developed and commercialized in the last decade, such as Pathwork Tissue of Origin Test® (Pathwork Diagnostics), with a potential application in tissue of origin prediction of metastases of known origin [87-89]. In these microarray-based studies the prediction accuracy ranged between 85-95%, both using frozen and FFPE samples, with poorly differentiated and undifferentiated primary tumors responsible for poor accuracy rates.

GEP prediction accuracy was compared in a blinded-multicentric study with standard IHC for the identification of the primary site in metastatic cancers of known origin. When a single round of IHC stains was sufficient to identify the primary site, IHC and GEP performance were similar, showing an accuracy higher than 90%; on the other hand, when a second round of stains was necessary, GEP proved to be superior than IHC (91% vs. 71%; p=0.023) in the precise prediction of the tissue of origin [90].

When applied to CUPs in retrospective and prospective settings, GEP studies reported a consistency with clinical evaluation and pathological assessment that ranged between 62.5-85% [91-94] as reported in **Table 3**.

Microarray is a complex, time-consuming and expensive technology, so, several research groups evaluated RT-qPCR as a cheaper, easier-to-use and rapid technology. Nevertheless, microarray data were useful to discover focused signatures and narrow the number of target genes able to discriminate between different tumor types.

An example is represented by the 92-gene signature identified by Ma et al. able to distinguish and classify 26 cancer types [95] with an overall sensitivity of 87% for tumor types and 82% for subtypes [96].

This assay, commercially distributed as CancerTypeID® (bioTheranostics), is a diagnostic tool employed in United States to solve uncertain cases that proved higher prediction accuracy when compared with standard IHC (79% v 69%) [97,98].

In addition, another signature of 10 genes was used to design a qPCR assay by Veridex, able to efficiently distinguish six cancer types (lung, breast, colon, ovary, pancreas, and prostate) [99].

Notably, applied to CUP, these assays predicted a tissue-of-origin that was consistent with the primary site unmasked a month-to-year after the initial diagnosis and/or clinicopathological features in 54-86% of cases (**Table 3**) [99-101][98].

Recently, a machine learning method called SCOPE (Supervised Cancer Origin Prediction Using Expression) was proposed as a diagnostic tool able to track down the tissue-of-origin using whole-transcriptome RNA sequencing data [102]. They used as reference set public datasets, including The Cancer Genome Atlas (TCGA) (N=10688 samples of 40 untreated primary tumor types and 26 adjacent normal tissues), while the testing was performed retrospectively on untreated primary mesothelioma (N=211) and treatment-resistant metastatic cancers (N=201). In this study, SCOPE achieved an overall accuracy of 99% on mesotheliomas and 86% for sarcomatoid carcinomas, and an accordance of 80% with conventional pathology in the 15 CUPs tested.

Only few clinical trials and case reports assessed the impact of GEP profiling on CUP patients survival when treated with site-specific regimens according to molecular prediction [98,101,103-105].

In a retrospective study, among 42 CUP patients predicted as CRC with high probability, 76% were treated with CRC-oriented regimens and had a median survival of 27 months, which was significantly longer than historical reports (8-11 months) [98]. In a similar study, 7 over 25 GEP-profiled CUPs were treated with

tumor-specific therapy according to the molecularly predicted primary site; of note, they showed a 5-year PFS of 25-72 months [103].

A large not-randomized study reported the outcome of 194 CUP patients treated according to molecular prediction [98]; of note, the median survival of the predicted tumors known to be responsive to therapy was significantly longer than resistant (13.4 v 7.6 months; p = .04).

Similar evidence was obtained from a prospective phase II clinical trial (NCT00936702) in which 46 CUP patients were enrolled and treated with carboplatin and paclitaxel [93]. For 38 patients an indication of a primary site was obtained using ResponseDX Tissue of Origin Test [88]: 19 of them were predicted as tumors sensitive to platinum/taxane therapy, whereas the other half included tumors in which this regimen is not the standard of care. The study assessed the response to platinum/taxane of these two groups and reported a significantly longer PFS (6.4 v 3.5 months; p= .026) and OS (17.8 v 8.3 months; p= .0052) in sensitive tumors.

Other studies provided data that support the inability of a site-specific therapy guided by molecular profiling to improve CUP patients' life-expectancy.

An example is represented by a randomized prospective phase II trial [106], that reported a 1-year survival rate lower in patients treated according to molecular predictions compared with broad-spectrum chemotherapy (44% v 54.9%; p .264) and no significant improvement in OS and PFS. However, in this case, primary site prediction was based on a poorly validated molecular test [107], thus is fundamental to careful monitor the assay performance.

More recently, similar results were also reported by a European phase III clinical trial (GEFCAPI 04, NCT01540058) that involved 243 CUP patients who underwent treatment with empiric chemotherapy combination (cisplatin and gemcitabine) or molecular GEP-based therapies (predictions obtained using Pathwork Tissue of Origin Test or CancerTYPE ID) [108].

Results from the phase III clinical trial NCT03278600 could potentially clarify the actual impact of site-of-origin profiling assays in predicting primary site and guiding therapy in CUP patients. However, despite the achievement of a clinical benefit still needs to be determined, the most recent NCCN CUP Guidelines [109] support the use of gene expression profiling to improve the diagnostic work-up of CUP patients.

Table 3. GEP-based tissue-of-origin profiling assays applied on CUP patients [110].

Study year and reference	Study type	Data type	N° of CUP patients profiled / enrolled	Type of sample	Method analysis	N° of tumor types	Tumor type included in the reference set	Most predicted tumor types	Validation	CUP prediction accuracy	Impact on clinical outcomes with prediction-based treatment
2005, [111]	Retrospective	mRNA	13	FF and FFPE	10500 -gene GEM	14	Breast, CRC, gastric, lung, melanoma, mesothelioma, ovarian, pancreas, prostate, renal, cSCC, head and neck, testicular, uterine	ND	Clinicopathological features	84%	ND
2006, [99]	Retrospective	mRNA	48	FFPE	10-gene RT-qPCR;	6	Lung, breast, colon, ovary, pancreas, and prostate	ND	Unmasked latent primary tumor sites identified months to years later from the initial diagnosis	76%	ND
2008,[92]	Retrospective	mRNA	38/38	FFPE	1900-gene GEM; CupPrint® (Agendia)	49	ND	Lung (24%), CRC (18%), pancreas (16%), ovarian (11%)	Clinicopathological features and IHC	85%	ND
2008, [101]	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
2008, [112]	Retrospective	mRNA	21	FFPE	1900-gene GEM; CupPrint® (Agendia)	49	ND	Colon (19%), small and large bowel (19%), ovarian (19%), breast (9%)	Clinicopathological features	67%	ND

2010, [91]	Retrospective	mRNA	16/21	FF	1550-gene GEM; Pathwork®TOO Test (Pathwork Diagnostic)	15	Bladder, breast, colorectal, gastric, hepatocellular, kidney, non–small cell lung, ovarian, pancreatic, prostate, thyroid carcinomas, melanoma, testicular, germ cell tumor, non-Hodgkin's lymphoma, and sarcoma	CRC (5), breast (4), ovary (3), lung (2), and pancreas (2)	Clinicopathological features and IHC	62.5%	ND
2011, [113]	Retrospective	miRNA	16	FFPE	MicroArray; 47-miRNA signature	10	Breast, colon, endometrium, stomach/gastric, kidney, liver, lung, pancreas, prostate, and melanoma	Gastric (31%), lung (25%), pancreas (19%), liver (12%), colon (6%), and kidney (6%)	Clinicopathological features	85%	ND
2011, [114]	Retrospective	miRNA	57/60	FFPE	48-miRNA RT-qPCR; miRNA	25	ND	ND	Clinicopathological features	ND	ND
2011, [115]	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
2012, [116]	Retrospective	DNA methylation	42	FF	DNA methylation array, 1505 CpG sites selected from 807 genes; GoldenGate Methylation Cancer Panel I (Illumina)	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
2012, [98]	Retrospective	mRNA	42	FFPE	92-gene RT-qPCR CancerTypeID® (bioTheranostics)	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 predictive tumor types

2013, [100]	Prospective/ Retrospective	mRNA	144/149	FFPE	92-gene RT-qPCR; CancerTypeID® (bioTheranostics)	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 tumor sites identified months to years later from the initial diagnosis	74-77%	ND
2013, [98]	Prospective	mRNA	252/289	FFPE	92-gene RT-qPCR CancerTypeID® (bioTheranostics)	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 controls
2013, [117]	Retrospective	miRNA	84/93	FFPE	64-miRNA RT-qPCR; miRviews met2	42	ND	ND	Clinicopathological features and IHC	92%	ND
2015, [103]	Retrospective	mRNA	30/25	FFPE	92-gene RT-qPCR; CancerTypeID® (bioTheranostics)	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	Carcinoma (40%) (of which 30% were germ cell tumors and 20% were neuroendocrine tumors), sarcoma (30%), melanoma (20%) and lymphoma (8%)	Clinicopathological features and additional IHC or genetic testing post molecular profiling	84%	Improved PFS compared to historical controls
2015, [94]	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

2016, [118]	Retrospective	DNA methylation	216	FF	DNA methylation array, 485 577 CpG sites; EPICUP assay	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
2016, [93]	Prospective phase II	mRNA	38/46	FFPE	2000-gene GEM; Pathwork® OO Test (Pathwork Diagnostic)	15	Bladder, breast, colorectal, gastric, hepatocellular, kidney, non–small cell lung, ovarian, pancreatic, prostate, thyroid carcinomas, melanoma, testicular, germ cell tumor, non-Hodgkin's lymphoma, and sarcoma	NSCLC (21%), CRC (18%), ovary (18%), pancreas (16%)	ND	ND	Improved OS in platinum-responsive tumor types
2016, [93]	prospective phase II	mRNA	38/46	FFPE	2000-gene GEM; ResponseDX Tissue of Origin™ Test (Cancer Genetics)	15	Bladder, breast, colorectal, gastric, hepatocellular, kidney, non–small cell lung, ovarian, pancreatic, prostate, thyroid carcinomas, melanoma, testicular germ cell tumor, non-Hodgkin's lymphoma, and sarcoma.	NSCLC (21%); CRC (18%); ovary (18%); pancreas (16%	ND	ND	Improved OS and PFS in platinum-responsive tumor types
2019, [102]	Retrospective	mRNA	12/15	FF and FFPE	Whole Transcriptome	40	ND	ND	Clinicopathological features	ND	ND
2019, [106]	Prospective phase II	mRNA	130	FF	22000-gene GEM; mRNA	39	ND	Pancreas (20.8%), gastric (20.8%), lymphoma (20%), urothelium (6.2%), cervix (5.4%), ovary (4.6%), biliary tract (3.8%)	ND	ND	No significant difference in PFS and OS of site-specific therapy compared to standard chemotherapy; prognostic value of the predicted sites-of-origin

Abbreviations: ND – not determined, GEM – gene expression microarray, FF – fresh frozen, FFPE – formalin-fixed paraffin-embedded tissue, OS – overall survival, PFS – progression-free survival, CRC – colorectal cancer, NSCLC – non-small cell lung cancer, HCC – hepatocellular carcinoma, ALL - Acute lymphoblastic leukemia, AML - Acute myeloid leukemia, CLL - Chronic lymphocytic leukemia, DLBCL - Diffuse large B-cell lymphoma, FL - Follicular Lymphoma, MCL - Mantle cell lymphoma, mBL - Monoclonal B-cell lymphocytosis, MM - Multiple Myeloma, MDS/MPS - Myelodysplastic Syndrome/Myeloproliferative Syndrome, cSCC - cutaneous squamous cell carcinoma.

Despite the high accuracy rates achieved by GEP assays, these tools have to face critical technical problems caused by the low-quality RNA derived from FFPE samples [119,120]. Formalin fixation followed by embedding in paraffin is still the standard technique to preserve tissue samples; for this reason, FFPE tissue samples are the most commonly available source of specimens. However, fixation and embedding processes significantly affect RNA integrity and can introduce nucleotide chemical modifications [121,122]. Moreover, other factors can also play a role to define RNA quality in FFPE samples: the time passed before the fixation, the tissue thickness, the temperature during fixation and the storage time of FFPE samples [122]. For these reasons, in most GEP-based studies, summarized in **Table 3**, technical analysis failures caused by the low-quality of RNA obtained occurred in a variable portion of the enrolled patients (i.e. ~15% using CancerTypeID assay [98]).

Predictive assays based on microRNAs

microRNAs (miRNA) expression was evaluated as a suitable alternative to gene expression to develop tissue-of-origin predictive assays because they are highly stable and resistant to RNAse degradation and can be robustly detected irrespective of the quality of the sample of archival specimen [123-126]. Molecular miRNA profiling of FFPE samples could be successfully obtained from all the available samples [127,128].

Given the key role of deregulated miRNAs in cancer, it is well-documented that the expression of specific miRNA signatures can efficiently discriminate tumor from normal tissues and different cancer types [129-132].

Similarly to GEP, microarray-based studies explored the feasibility of a miRNA-based detection of the tissue-of-origin and reported accuracy rates that ranges between 78-90% when dealing with metastatic cancers of known origin [113,133,134].

Rosenfeld et al. assessed the miRNA profiles of 253 samples (up to 25 different histological subtypes) and built two different classifiers by using decision-tree and K-nearest neighbors (KNN) algorithms [134]. They identified a 48-miRNA signature that showed high accuracy to determine the appropriate tissue-of-origin of metastases of known origin. Later, this molecular assay was translated from microarray to RT-qPCR technology, maintaining its analytical validity [134,135] to

finally become the miRview mets assay, which is an LDT in USA distributed by Rosetta Genomics.

When applied to a prospective series of 74 CUPs (**Table 3**), molecular prediction was found in agreement with clinicopathological features in 84% of cases [115]. In addition, in another study [114], this assay was also used to predict the site-of-origin of 57 retrospective CUP patients that metastasized in the central nervous system; of note, the prediction's outcome matched in 88% of cases with clinicopathological hypothesis at the diagnosis, information obtained later during the course of the disease or specific examinations performed after and/or based on the molecular prediction.

Rosetta Genomics developed also an updated second-generation version of this assay, named miRviews met^2 (Rosetta Cancer Origin Test), that allowed primary prediction among 42 different tumor classes. Custom microarray data on a training set of 1282 samples were used to build a classifier with a combination of two different algorithms (decision-tree and KNN) that allowed tumor classification based on a 64-microRNAs signature. When applied on an independent blinded set of 509 primary tumors and metastases of known origin, the accuracy of this tool was 85% [128]. On the contrary, this analysis performed on 84 CUP samples reported a molecular prediction that showed accordance in 70% of patients with the first clinical hypothesis, in 89% with treatment response and in 92% with the final clinical diagnosis (when achieved) according to supplemental IHC stains of patients [117].

Moreover, another microarray-based molecular study [113] identified a 47-miRNA signature able to distinguish ten different cancer types with a prediction accuracy of 100% in primary cancers and 78% in metastases of known origin, which outperformed Rosenfeld's signature when applied to the same dataset. This signature was also used to predict the primary site of 16 CUPs (**Table 3**), showing high consistency with clinical and pathological hypotheses.

The employment of these assays was also reported in few case reports that proved their potential clinical utility. One example is represented by a female CUP patient with a suspected peritoneal or ovarian cancer, whose molecular prediction by miRviews met^2 indicated a mesothelioma origin; in this case, additional IHC stains confirmed the diagnosis, leading to a therapeutic switch that resulted in a significant improvement of patient's life-expectancy [136]. Another study regarded a male CUP patient that was predicted with a probability of 90% as breast cancer by miRviews

met[2] that was later confirmed by pathologic evaluation and IHC on breast resected metastasis, thus allowing optimal treatment for the patient [137].

Predictive assays based on DNA methylation

DNA methylation in CpG sites, as regulatory mechanism of transcription, can also be regulated in cancer and play an important role in cancer development and progression. A DNA methylation fingerprint of 1628 human samples was obtained interrogating 1505 CpG sites using Illumina GoldenGate Assay [116]. Specifically, Fernandez et al. analyzed 424 normal tissues, 150 non-cancerous disorders and 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. Of note, this study highlighted the existence of tissue-type specific methylation patterns and the impact of methylation in tumorigenesis; specifically, human cancer cells progressively undergo to hypermethylation of CpG in CpG-island promoters and, concurrently, to hypomethylation of CpG in non-CpG-island promoters. Moreover, DNA methylation fingerprints of CUPs were compared with the analyzed primary and metastatic tumors with known origin and a site-of-origin was assigned in 69% of CUP patients, fully confirmed by pathological information in 78% of cases.

More recently, Moran et al. [118] developed a classifier based on microarray DNA methylation signatures (EPICUP assay). This study included a reference set of 2790 known primary tumors, belonging to 38 tumor classes and including 85 metastatic tumors. When applied to 216 CUP samples, the molecular prediction was in accordance in 90% of cases with the primary site revealed during the disease follow-up or post-mortem autopsy or available pathological assessment. Moreover, since survival and therapy information of 114 among the 216 CUP patients was available, they were able to observe a significant improvement in OS of patients who received a tumor type-specific therapy (13.65 months) compared to those who received empiric treatments (6 months) [118].

In a recent in-silico study [138], *Tang et al*. proposed a miRNA- and methylation-based classifier that proved to be more efficient in the identification of predictive features (miRNAs or methylation level) and thus in primary site prediction. For this purpose, they used The Cancer Genome Atlas (TCGA) data to obtain miRNA profiles from 6602 samples and DNA methylation profiles from 5379 samples, as training set that comprehended 14 tumor classes. Then, they evaluated the

performance of these classifiers on independent datasets, showing an overall accuracy of 95% for DNA methylation and 91% for miRNA. However, this approach has not been tested on CUPs yet.

Driver mutations and precision medicine

Different studies reported a mutational scenario characterized by high heterogeneity across different CUP patients [139-141] (summarized in **Table 4**); of note, mutations, for the most part, harbored in common cancer driver genes, mostly involved in core mitogenic, cell growth pathways (especially PI3K and MAPK pathways) and epigenetic deregulation [142-144].

In a small portion of CUP patients it was possible to recognize a common molecular signature, mainly related to exogenous mutagens exposure, such as smoking or UV radiation [141,142].

Stratifying patients according to CUP histological subtypes, it was possible to observe that adenocarcinoma and poorly differentiated carcinoma subtypes had mutations in genes involved in signal transduction pathways, while squamous cell carcinomas were mutated, for the most part, in cell cycle control and DNA repair genes [140]. Alterations in the receptor tyrosine kinase (RTK)/Ras signaling pathway were observed prevalently in CUPs belonging to adenocarcinoma subtype compared to non-adenocarcinomas (72% v 39%, p< .001) [139]. Moreover, adenocarcinoma histology was also predominantly characterized by amplification or mutation of *ERBB2* (10% v 4%) and alterations of *EGFR* (8% v 3%) or *BRAF* (6% v 4%) [139]. *MET* was found to harbor mutations more frequently in CUP patients compared to metastases of known origin, suggesting its potential role to enhance CUP aggressiveness [145]. Of note, *MET* and *CTNNB1* alterations were found to correlate with a worse survival (OS 11 v 21 months, p=.015) [146].

In addition, chromosomal instability was recently hypothesized as a driver cause of the unfavorable CUP subset [37].

With the advent of precision medicine, we are rapidly moving forward a more precise, predictable and powerful healthcare that is customized for each individual cancer patient, regardless to the primary site. Such approach, combined with the availability of public multi-omics databases projects, such as The Cancer Genome Atlas (TCGA) [147], is revolutionizing cancer classification, that relies more and more on molecular information than histological one. Personalized medicine is

shaping the future of cancer therapy [148,149] and consists in the management of cancer patients in relation with the discovery of specific molecular features, based on which the best therapeutical option is selected (**Figure 5**).

Results from the IMPACT study (Initiative for Molecular Profiling and Advanced Cancer Therapy), in which patients with advanced cancers received matched targeted therapies (MTT) based on their mutational profile, strongly support the employment of genomic matching. Specifically, 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, patients treated with MTT and respond to therapy had a significantly longer OS compared to nonresponders (23.4 v 8.5 months; p < .001) [150].

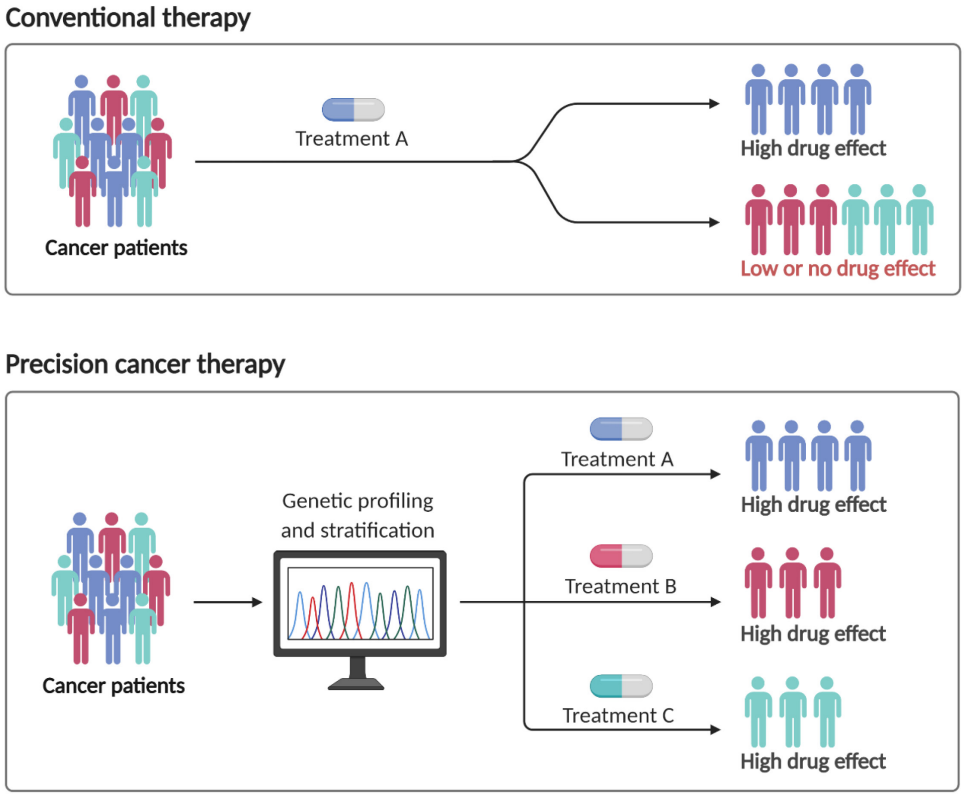

Figure 5. Conventional vs Precision Medicine. Precision medicine is based on molecular data collection (i.e. genetic, epigenetic and transcriptomic data) that could potentially improve patients' stratification and the identification of the best treatment option for each subgroup (reprinted from "Precision Cancer Therapy", by BioRender.com).

The identification of druggable alterations in CUP patients could potentially improve their treatment options and positively affect their survival. Based on this assumption, several studies analyzed the mutational landscape of CUP patients [140,141,151].

Ross et al., in a comprehensive retrospective study, used the 236-gene FoundationOne assay (Roche) to explore the genomic profiles of 200 CUPs [151]. At least 1 clinically relevant genetic alteration was found in 96% of CUPs, with a mean of 4.2 alterations per tumor. The most frequently mutated genes were *TP53* (55%), *KRAS* (20%), *CDKN2A* (19%), *MYC* (12%), *ARID1A* (11%) and *MCL1* (10%).

Recently, at Memorial Sloan Kettering Cancer Center, a comprehensive 468-gene panel called "MSK-IMPACT" was designed and used to analyze up to 10,000 tumor samples, including 186 CUP patients [152].

Mutational data from these patients is publicly available at cBioPortal [153], and reported at least 1 mutation in 92% of CUPs. The most common mutations harbored in *TP53* (45.2%), *KRAS* (22.6%), *ARID1A* (14.5%), *SMARCA4* (12.9%), *KEAP1* (12.4%), *KMT2D* (9.7%), *CREBBP* (9.7%), *PIK3CA* (9.1%), *STK11* (8.6%) and *TERT* (8.6%); while the most common CNA were homozygous deletions of *CDKN2A* (10.2%) and *CDKN2B* (9.7%) and amplifications of *CCNE1* (4.3%), *ERBB2* (4.3%) and *MCL1* (4.3%) [152]. Moreover, the MSK-IMPACT study reported alterations on *TERT* promoter, which are also associated with a poor prognosis [154,155].

Another consulting database named "GENIE" [156], was developed by AACR (American Association of Cancer Research) and collects whole exome sequencing mutational data from more than 120,000 cancer samples, including 3560 CUPs (*last update 31/07/2021*). CUP samples has at least 1 mutation in cancer-related genes and the most common alterations were in *TP53* (45.2%), *KRAS* (18.8%), *KMT2D* (13.8%), *ARID1A* (11.1%), *PIK3CA* (9.4%) and *APC* (8.1%); homozygous deletions of *CDKN2A* (11.9%) and *CDKN2B* (11.7%) and amplifications of *MYC* (4.8%) were the most frequent CNA [156]. According to *Ross et al.*, potentially actionable mutations were detected in about 20% of CUPs. Other studies reported rates that ranged from 15 to 96% (reported in **Table 4**) that made necessary to define more stringently the actionability criteria for mutation selection.

With this purpose, *Varghese et al*. [141] analyzed the frequency of druggable genetic alterations in 150 CUP samples tested for MSK-IMPACT gene panel using OncoKB resources (http://oncokb.org, [157]), a knowledge database that contains

information about the impact of somatic alterations on function and their treatment implications. In this study, the authors were able to identify potentially druggable alterations in 30% of CUP cases (FDA level 2-3 of evidence for actionability). Applying the same selection criteria to the dataset of *Ross et al*. [151], frequencies of CUP cases potentially eligible for basket trials were found to be consistent.

The utility of biological target therapy on CUP patients is considered controversial in the scientific community; in fact, response rate to these agents varies a lot across different cancer types, making more challenging the selection of the right treatment in a tumor-agnostic setting. In addition, specific driver mutations could be predictive of response in some tumor types but not others. One example is represented by *BRAF* mutation (V600E), which is associated with a good response to BRAF inhibitors in melanoma but not in colorectal cancer [158].

In the SHIVA study, a prospective, randomized phase II clinical trial, target therapy was administered in a primary-blinded setting to patients with metastatic solid tumors, including 5 CUPs; no significant improvement was observed in these patients [159]. Results from the MOSCATO (Molecular Screening for Cancer Treatment Optimization) and ProfiLER trials reported that about 30% of genetically profiled tumors were suitable for targeted therapy, but only 6-7% of them was actually treated with this approach [160,161].

Genomic profiling of CUPs demonstrated a meaningful clinical relevance in several studies and case reports supporting treatment selection and resulting in an improvement in patient's survival [104,162-166].

In a phase I trial, among 17 CUP patients mutationally profiled, 7 were suitable for specific matched targeted therapies; of note, 3 patients had stable disease and 1 a mixed response [144].

Despite these promising results, basket trials on metastatic solid tumors, including CUP patients, poorly succeeded to demonstrate the utility of target therapy in CUP. *Hainsworth et al.* proposed the combination of molecular and genomic profiling as a potential strategy to guide treatment decision [104]. In this study, among 21 molecularly-profiled CUP patients (miRNA 92-gene molecular cancer classifier assay, CancerTYPE ID) predicted as NSCLC, 4 were found to harbor EML4-ALK fusion genes and 1 patient was treated with ALK inhibitor and had a prolonged benefit. Similarly, results from a prospective study called SUPER (Solving Unknown

Primary Cancer) involving 172 CUP patients supported the employment of a joined approach using mutational and GEP profiling to improve CUPs management [167]. Results from ongoing clinical trials or basket trials extended to CUP patients could help clarifying the role of driver mutations in their clinical management. In particular, in the ongoing clinical trial of Foundation Medicine (NCT02628379) CUP patients are treated according to their specific druggable alterations determined by FoundationOne CDx next generation sequencing (NGS) assay, irrespective of their primary site.

Moreover, recent studies evidenced how CUP patients could benefit from treatment with immune checkpoint inhibitors [168]. Researchers are intensely searching for novel biomarkers of response to immunotherapy, that to date include programmed cell death ligand 1 (PD-L1) expression, high tumor mutational burden (TMB) and high microsatellite instability (MSI); of note, high MSI (MSI-H) has been reported to occur in 28% of CUP patients [143,169,170].

NCCN guidelines suggest the use of pembrolizumab, an FDA-approved anti-PD1 antibody, on MSI-H or with deficiency of the mismatch repair system (dMMR) patients regardless of the primary site [48]; in fact, this immune checkpoint inhibitor proved its antitumor activity across different tumor types, including colorectal cancer, melanoma and non-small cell lung cancer [171-174]. However, CUPs with MSI-H or dMMR represent a small portion [169]. To our knowledge two studies reported CUP cases (1 with high PDL1 expression and 1 with dMMR) that received immunotherapy treatment with promising outcome [162,175].

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). or positively concluded (UMIN000030649). In this trial, 56 CUP patients, of whom 45 previously treated, were treated with anti PD-1 nivolumab; the objective response rate (ORR) in the overall population was 21.4%, with a median PFS and OS of 5.1 (95% CI, 2.7-5.6) and 15.9 months (95%CI, 8.4-not reached), respectively [176].

Results from the ongoing global phase II clinical trial of La Roche-Foundation Medicine (NCT03498521) could also help clarify the impact on OS and PFS of novel targeted therapy agents and immunotherapy on unfavorable CUP patients compared to chemotherapy [177].

Table 4. Studies analyzing the mutational landscape of CUP tumors [110].

Study	Analyzed genes	reference	Year	N° of CUP patients	% of samples with at least one alteration	% potentially targetable alterations	Most common genetic aberrations
Retrospective target sequencing and CNA analysis	701	[142]	2013	16	100	81	TP53 (62%), GNAS (25%), NOTCH1 (18%), NOTCH2 (18%), CDKN2A (18%), PIK3CA (18%), BRCA1 (18%), STK11 (18%).
Retrospective metanalysis of mutational profiling, CNA analysis and protein expression profiling	47	[143]	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	[146]	2014	87	66	37	CTNNB1 (19.5 %), KRAS (10.2%), PIK3CA (7%), BRAF (4.5%), MET (4.5%).
Retrospective target sequencing using FoundationOne assay	236	[139]	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%), BRCA2 (6%).
Retrospective target sequencing using Oncomine Focus Assay and Cancer Hotspot v2 panel	76	[140]	2018	21	81	52	TP53 (47%), KRAS (12%), MET (12%) and MYC (12%).
Retrospective target sequencing using MSK impact assay	468	[141]	2017	150	91	30	TP53 (70%), KRAS (35%), CDKN2A (30%), KEAP1 (23%), SMARCA4 (22%)
Retrospective targeted sequencing and copy number analysis	50	[178]	2016	55	84	15	TP53 (55%), CDKN2A (22%), KRAS (18%), SMAD4 (11%), FGFR3 (9%), ATM (7%), RB1 (7%).
Retrospective target sequencing using NGS CLIA-certified assay	-	[144]	2017	17	88	41	Impaired P signaling (47%), Epigenetic deregulation (47%), impaired cell cycle control (47%)
Retrospective target sequencing using NGS CLIA-certified assay	592	[169]	2018	389	ND	22	TP53 (54%), KRAS (22%), ARID1A (13%), PIK3CA (9%), CDKN2A (8%), SMARCA4 (7%)

Abbreviations: CNA- copy number alteration, OS – overall survival, PFS – progression free survival, RFS – recurrence free survival, N/E – not evaluated, no multivariate analysis conducted

Mutational landscape for primary site prediction

Riding the wave of tissue-gnostic approach, proof-of-concept classifiers were developed also using CUP genomic profiling data. *Marquard et al.* [179] used publicly available databases to build an accurate classifier, called TumorTracer [180], based on point mutations and copy number alterations of 232 cancer-related genes and 96 classes of base substitutions, and suggested its potential utility for primary site detection in CUP patients.

The PCAWG (Pancancer Analysis of Whole Genomes) consortium developed a deep learning based classifier of the primary site based on the analysis of somatic passenger mutations detected by whole genome sequencing of 2606 tumors, belonging to 24 cancer types[181]. Such approach reached an accuracy of 88% and 83% when applied on independent cohorts of primary and metastatic tumors, respectively.

A hint of the potential of a combined approach on CUPs can be found in recently published case report [182]. In this study, a 67-year-old male patient got, after three misdiagnoses, a final diagnosis of malignant melanoma thanks to the combined application of NGS analysis, a patient-derived xenograft model and tissue-of-origin prediction using TCGA transcriptomic data.

Penson et al. [183] applied a machine-learning approach for tissue-of-origin prediction using target DNA sequencing data. The study comprehended a training set of 7,791 patients of advanced tumors belonging to 22 different classes and a test set of 11,644 cases. Prediction accuracy observed in the two cohorts was ~74% and remain similar when tested on plasma cell-free DNA mutations, suggesting a great potential of liquid biopsy for cancer diagnosis and primary site identification. When applied on 141 CUP patients, Penson's algorithm allowed an indication of a likely primary site in 67.4% cases. Moreover, such approach, applied on two prospective patients with a first hypothesis of metastatic breast cancer, lead to a final diagnosis and a change of treatment with an evident clinical benefit for these patients.

Liquid biopsy for primary site prediction or treatment decision

Liquid biopsy, using circulating tumor cells (CTCs) or circulating tumor DNA (ctDNA), opens a new window of opportunity in the clinical management of CUP patients and allows the collection of genetic data from both the primary and metastatic sites, with a less invasive approach (**Figure 6**). Indeed, tissue samples

are usually obtained at the time of diagnosis by standard tissue biopsy, when feasible. It is not representative of the whole tumor heterogeneity [184] and can be altered by formalin fixation process [185].

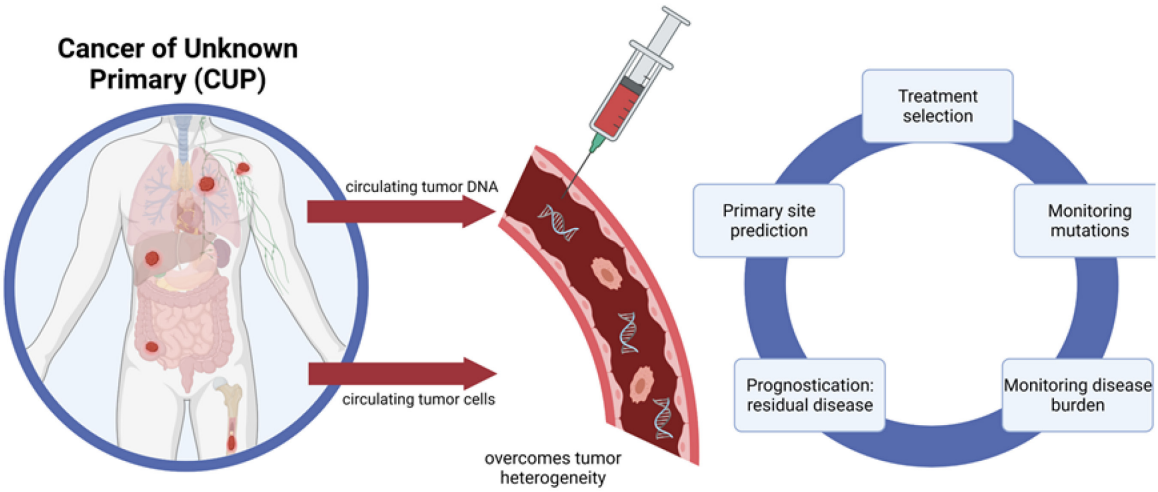

Figure 6. Potential clinical application for liquid biopsy in cancers of unknown primary. Liquid biopsy, specifically focusing on ctDNA and CTCs, have promising applications in CUP. This analysis could be employed for primary site prediction, treatment selection, monitoring of the dynamic changes of specific mutation, prognosis and disease burden evaluation. Figure created with BioRender.com adapted from [186].

On the contrary, since CTCs and ctDNA are shed into the bloodstream from multiple metastatic sites, they can represent inter- and intra-tumor heterogeneity of sub-clonal populations and clearly provide a more realistic genetic information that could guide treatment selection. In addition, CUP genetic testing is currently limited by the scarcity of tumor biopsy material and reduced biopsy DNA quality. Moreover, since liquid biopsy requires a simple blood withdrawal, it can be performed serially during patient's follow-up, providing useful data for the evaluation of treatment response, monitoring the dynamic changes of specific mutations, prognosis and disease burden [31,186].

Several proof-of-concept studies explored on different tumor types (but not in CUPs), the utility of genetic and epigenetic profiling of ctDNA, CTCs staining, transcriptomic analysis of tumor-educated platelets (TEPs) or proteomic analysis of plasma-derived extracellular vesicles and particles (EVPs) aimed to cancer early detection or tissue-of-origin prediction [187-193]. In another proof-of-concept study, genome-wide DNA methylation data of a set of solid tumor samples (TCGA) and healthy plasma samples were used to select the CpG clusters and build a classifier,

named CancerLocator, able to successfully distinguish ccfDNA from healthy or cancer patients, and also provide an indication of the organ of origin [194]. The staining of CTCs with a set of five fluorophore-conjugated antibodies (CK7, CK20, TTF1, ER, PSA) was proposed as a strategy able to recognize the most common tumor types, namely lung, colon, breast and prostate cancers. However, when dealing with poorly differentiated tumor, this approach failed to reach a diagnosis [195].

Few studies included a limited number of CUP patients, so, to date available data is not sufficient to support liquid biopsy application in the clinical management of CUPs.

Different works supported circulating tumor DNA (ctDNA) and circulating tumor cell (CTC) testing for disease monitoring and treatment selection [196]. In addition, ctDNA variant allele frequency was found to be associated with a poor prognosis in patients with metastatic disease [197].

Several studies demonstrated the utility of a liquid biopsy -based genomic analysis using ccfDNA from cancer patients (and few CUPs), reporting high sensitivity rates in the identification of oncogenic and actionable alterations genomic data [198-200]. However, *Kato et al.* [175] were able to perform this analysis on a cohort of up to 442 CUP patients. Analyzing 54-70 genes, they detected at least 1 alteration in 80% of CUPs, with *TP53* (37.8%) as the most mutated gene followed by mutations in genes involved in MAPK pathway (31.2%), PI3K signaling (18.1%) and cell cycle (10.4%). In the same study, such approach was applied to monitor the dynamic changes of ctDNA during therapy of an 82-year-old male CUP patient. The detection of two mutations (*KRAS* G12D and *MLH1* R389W) led to a combined treatment with nivolumab and trametinib that resulted in patient's partial response after 8 weeks from the first administration [175].

CTCs and CTC clusters are highly metastatic [201] and their number in the blood correlates with patient prognosis in several cancer types [202-204]. Different strategies for size-based or antigen-based CTC isolation and/or enumeration have been explored in the past ten years, each one presenting advantages and limitations [205].

CTC count, obtained with the FDA-approved CellSearch System (Menarini-Silicon Biosystems), became a prognostic biomarker in breast cancer and other tumor types [206]. However, such approach requires the selection of specific tumor

antigens for CTC isolation, which is very challenging when dealing with poorly undifferentiated CUPs [207].

Beyond the count, CTCs have also potential application in *in vitro* [208] and *in vivo* disease modeling [209] and drug testing [210]. In metastatic patients it is possible to find approximately 10 CTCs/mL of blood [211].

Patients with CUP, which is the epitome of advanced metastatic disease, should have a high number of CTCs and CTC-clusters in the circulation. To date, only few works evaluated CTC count in CUP using CellSearch System or immunofluorescent methods reporting CTC presence in >50% of CUP patients [207]. Specifically, Komine et al. reported a median number of 31 CTCs in CUP patients before treatment that had a decreasing trend after chemotherapy, thus suggesting its potential as biomarker of response to therapy [207].

Aim

The first aim of this study is the development and application of a microRNA (miRNA)-based molecular profiling assay able to predict the tissue-of-origin of metastatic tumors, including Cancers of Unknown Primary (CUPs), and potentially offer valuable indication about the possible primary site, thereby supporting treatment decisions.

To fulfill this aim, we designed a miRNA-based molecular assay for the rapid, on-demand quantification of a pre-determined set of 89 miRNAs in formalin-fixed and paraffin-embedded (FFPE) clinical samples. Unlike previous assays, our test employs Droplet Digital PCR (ddPCR) technology to assess miRNA absolute level. Using this molecular tool, we analyzed 159 FFPE samples including a reference set of primary tumors, belonging to 17 tumor types and 19 histological subtypes, and a test set that comprises metastases of known origin and unknown/uncertain origin, with a 100% technical successful rate. This molecular tool, taking advantage of two predictive classifiers, succeeded in predicting the most probable primary tissue(s) of a large collection of up to 53 cancers of unknown or uncertain origin. Prediction outcome showed high concordance in multiple metastases from the same CUP patients and strong consistency with first pathological/clinical hypotheses, when available.

The second objective of this study is to explore, using two prospective CUP patients, the utility of liquid biopsy, specifically ctDNA and CTC-based applications, as non-invasive approaches to detect CUP druggable mutations. We compared two methods to isolate CTCs: one antigen-based, size-agnostic (CELLSEARCH) and another antigen-agnostic, size-based (Parsortix). CTCs and ctDNA were detectable in the blood of CUP patients and analyzed for genomic alterations, which were compared with those identified in tumor biopsy using different gene panels. When covered by all panels, the identified mutations were highly concordant, thus confirming the advantage of liquid biopsy genetic testing in CUPs, especially using CUP-dedicated NGS panels.

Both molecular diagnostics and genetic profiling, are extremely relevant to offer these patients novel therapeutic options and potentially improve their life-expectancy and should become the standard in CUP management.

Materials and Methods

Patients and samples

In this study we collected a total number of 159 formalin-fixed paraffin-embedded (FFPE) samples from 150 patients that received a diagnosis and were treated at Sant'Orsola-Malpighi Bologna University Hospital, Italy (N=84), at the University Hospital of Ferrara, Italy (N=52) or at the Medical University of Graz, Austria (N=14). The study cohort (summarized in **Table 5**) consisted of patients with tumors 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).

Primary tumors, used as training set, included samples obtained from the following tumor sites/types: lung (LUAD, adenocarcinoma, N=6 and LUSC, squamous cell carcinoma, N=3), pancreas (PAAD, exocrine adenocarcinoma, N=5), ovary (OV, ovarian serous carcinoma, N=6), liver (LIHC, hepatocellular carcinoma, N=6), biliary tract (CHOL, cholangiocarcinoma, N=6), kidney (KICA, which includes kidney renal clear cell carcinoma or KIRC, N=5 and kidney renal papillary cell carcinoma or KIRP, N=3), colorectum (CRC, adenocarcinoma, N=7), testis (TGSC, germ cell seminomatous carcinoma, N=4), endometrium (UCEC, adenocarcinoma, N=5), stomach (STAD, adenocarcinoma, N=5), bladder (BLCA, transitional cell carcinoma, N=4), breast (LBC, luminal non-special type and lobular breast carcinoma, N=5), triple negative breast cancer (TNBC, N=3), prostate (PRAD, adenocarcinoma, N=5), melanoma (SKCM, melanoma of skin, N=7), head and neck (HNSC, squamous cell carcinoma, N=6) and gastrointestinal neuroendocrine carcinoma (GI-NET, N=5). In addition, as test set, we analyzed 10 metastases of known origin that derived from lung, melanoma, stomach, prostate, head and neck, kidney, colon, breast, pancreas and endometrium. A final number of 53 CUP samples were included in this study, specifically 43 retrospective and 10 prospective cases. Of note, from 5 retrospective CUP patients we were able to obtain multiple biopsies collected from synchronous and/or metachronous metastases that were independently analyzed. CUP diagnosis was obtained after detailed clinical and pathological investigations as suggested by CUP diagnostic guidelines [5,48]. For the majority of samples, an extensive IHC panel was assessed at the time of diagnosis and the outcome was recorded (**Table 9**). However, it is worth mentioning 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 2005-2009, 26 between 2010-2014 (49%) and 13 between 2015-2019 (25%).

For each sample, we obtained 2-5 10μm-thick tissue sections. The first section was stained with hematoxylin-eosin (HE) and examined by an expert pathologist to select the tumor area, which was macrodissected prior RNA extraction. Tumor cell fractions were selected to have at least 30% cellularity.

From two prospective CUP patients included in our cohort (CB055 and CB090) and recruited at Bologna University Hospital we were able to collect also blood samples. For CB055, blood sampling was performed at three different time points: A) at diagnosis (August 2018), B) during FOLFOX-4 treatment (stable disease, November 2018) and C) at disease progression (May 2019). For Pt#95, blood sampling was performed at diagnosis.

Plasma separation was performed via centrifugation at 1900 x g for 10 minutes at 4°C. A variable number (N=2-5) of plasma aliquots (1 mL) for each patient was collected and stored at -80°C before the isolation of circulating cell-free DNA (ccfDNA). PBMC were isolated from peripheral blood of CB090 using Ficoll-Paque Plus (17-1440-02, GE Healthcare, Chicago, IL, USA). Briefly, after plasma depletion an equal volume of PBS was added to the remaining blood in EDTA tubes. Following, 4 ml of diluted blood was stratified on 3 mL of Ficoll-Paque Plus and centrifuged at 400 x g for 30 minutes at room temperature in a swinging-bucket rotor. We removed carefully the PBMC ring by pipetting, washed the pellet with PBS and centrifuged at 100 x g for 5 minutes. Finally, the pellet was stored at -80°C prior to proceed with DNA extraction.

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

RNA extraction and cDNA conversion

We isolated total RNA, including microRNAs, from the tumor FFPE sections using miRNeasy FFPE kit (Qiagen), performing deparaffinization with xylene and following the instruction from step 11 of the *Protocol: Purification of Total RNA, Including miRNA, from FFPE Tissue Sections*. 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). All samples were suitable for the molecular testing.

RNA conversion to cDNA was performed using the miRCURY LNA RT Kit (Qiagen). The 10 µL-reaction mix was prepared for each sample by mixing: 2 µL of 5x Reaction Buffer, 4.5 µL of nuclease free water, 1 µL of enzyme mix, 0.5 µL of UniSp6 RNA spike-in and 2 µL of diluted RNA (10 ng of total RNA). The resulting cDNA was stored in LoBind DNA Eppendorf tubes (Eppendorf) at -20°C. For each sample, a RT-qPCR was performed as quality control step using 5 µL SsoAdvanced Universal SYBR Green Supermix (BioRad) and 1 µL of miRCURY LNA miRNA PCR Assay (Qiagen) testing UniSp6 and SNORD44 targets. A 10 µL- reaction mix was set further adding 4 µL of 1:80 diluted cDNA of each sample. This step was crucial to evaluate RT reaction efficiency, according to UniSp6 threshold cycle (Ct), and RNA integrity and amplifiability, evaluating SNORD44. SNORD44 Ct was used to establish the appropriate cDNA dilution to be employed in digital droplet PCR (ddPCR) analysis. Specifically, when SNORD44 Ct ranged between 24-30 (threshold set at 160), cDNA was diluted 1:50; for Ct below 24, cDNA was diluted 1:100-1:200 and when Ct was higher than 30, the RT was repeated again using undiluted RNA and qPCR analysis repeated. In addition, cDNA had an additional 1:10 dilution in miR-21-5p and UniSP6 wells. Applying these criteria, we prevented ddPCR saturation problems or low miRNA expression levels in ddPCR analysis.

DNA extraction

For two prospective CUP cases we also obtained the DNA: specifically, from FFPE tissue biopsy of a lymph node for CB055 and the ampulla of Vater for CB090. The tumor area was selected by an expert pathologist in the hematoxylin-eosin stained section, used to guide the microdissection of the tissue from unstained slides. DNA was extracted using QIAamp DNA FFPE Tissue Kit (Qiagen) following manufacturer's guidelines. Normal DNA was obtained from PBMCs isolated from the peripheral blood of both patients. DNA from PBMCs was extracted using QIAamp DNA Micro kit (Qiagen). Ten slices of CB055 FFPE tumor tissue were sent to Roche Foundation Medicine (Cambridge, MA, USA) for comprehensive genomic profile analysis with FoundationOne CDx (F1CDx) assay.

Circulating cell free DNA (ccfDNA) was purified from 1 mL of plasma by Maxwell RSC instrument (Promega) using Maxwell RSC ccfDNA Plasma Kit (Promega). Genomic and circulating cell-free DNA were quantified using the Qubit 4.0

Fluorometer (ThermoFisher Scientific) and the Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific).

Yields and quality of FFPE-derived gDNAs was assessed using NGS FFPE QC Kit (Agilent Technologies) according to the producer's protocol (G9700-90000, v. E1, May 2018, Agilent Technologies).

MicroRNA selection

The miRNA-signature for tumor primary site prediction used in this study was obtained by integrating published signatures of two microarray-based molecular studies [127,134] and adding 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) as detailed in **Table 5**. Specifically, Rosenfeld et al. [134] analyzed up to 25 different histological subtypes and discovered 48 miRNAs that were able to efficiently discriminate the primary tumor in metastases of known origin; similarly, in the second study [113], using a training set that included 10 tumor classes, Ferracin et al. identified a 47-miRNA signature that was successfully applied to test metastases of known and unknown origin. The additional miRNAs we decided to include in the panel were selected as candidate reference miRNAs or to widen the number of the assessed miRNAs, 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 5. List of miRNA assays in the custom ddPCR plate miRNA used by PAMr and LASSO for the site-of-origin prediction [212].

microRNA	Plate well	Target sequence	Qiagen LNA assay	Reference	used in PAMr prediction	used in LASSO prediction
hsa-let-7e-5p	A01	UGAGGUAGGAGGUUGUAUAGUU	YP00205711	[134]	**x**	**x**
hsa-let-7i-5p	A02	UGAGGUAGUAGUUUGUGCUGUU	YP00204394	[134]	**x**	**x**
hsa-miR-103a-3p	H12	AGCAGCAUUGUACAGGGCUAUGA	YP00204063	[213]	**x**	**x**
hsa-miR-106b-5p	A06	UAAAGUGCUGACAGUGCAGAU	YP00205884	[134]	**x**	**x**
hsa-miR-10a-3p	A04	CAAAUUCGUAUCUAGGGGAAUA	YP00205688	[127]	**x**	
hsa-miR-10a-5p	A03	UACCCUGUAGAUCCGAAUUUGUG	YP00204778	[127]	**x**	**x**
hsa-miR-10b-5p	A05	UACCCUGUAGAACCGAAUUUGUG	YP00205637	[134]	**x**	**x**
hsa-miR-122-5p	A07	UGGAGUGUGACAAUGGUGUUUG	YP00205664	[127]		
hsa-miR-124-3p	A08	UAAGGCACGCGGUGAAUGCCAA	YP00206026	[134]	**x**	
hsa-miR-126-5p	A09	CAUUAUUACUUUUGGUACGCG	YP00206010	[127]	**x**	

hsa-miR-130a-3p	A10	CAGUGCAAUGUUAAAAGGGCAU	YP00204658	[134]	x	x
hsa-miR-135b-5p	A11	UAUGGCUUUUCAUUCCUAUGUGA	YP00204130	[127]	x	x
hsa-miR-138-5p	B01	AGCUGGUGUUGUGAAUCAGGCCG	YP00206078	[134]	x	
hsa-miR-141-3p	B02	UAACACUGUCUGGUAAAGAUGG	YP00204504	[127]	x	x
hsa-miR-142-3p	B03	UGUAGUGUUUCCUACUUUAUGGA	YP00204291	[134]	x	x
hsa-miR-145-5p	B04	GUCCAGUUUUCCCAGGAAUCCCU	YP00204483	[127]	x	x
hsa-miR-146a-5p	B05	UGAGAACUGAAUUCCAUGGGUU	YP00204688	[127]	x	
hsa-miR-148b-3p	B06	UCAGUGCAUCACAGAACUUUGU	YP00203906	[134]	x	
hsa-miR-149-5p	B07	UCUGGCUCCGUGUCUUCACUCCC	YP00204321	[127]	x	x
hsa-miR-152-3p	B08	UCAGUGCAUGACAGAACUUGG	YP00204294	[134]	x	
hsa-miR-16-5p	H06	UAGCAGCACGUAAAUAUUGGCG	YP00205702		x	x
hsa-miR-181a-2-3p	B10	ACCACUGACCGUUGACUGUACC	YP00204142	[127]	x	
hsa-miR-181a-5p	B09	AACAUUCAACGCUGUCGGUGAGU	YP00206081	[134]	x	x
hsa-miR-181b-5p	B11	AACAUUCAUUGCUGUCGGUGGGU	YP00204530	[134]	x	
hsa-miR-182-5p	C01	UUUGGCAAUGGUAGAACUCACACU	YP00206070	[127]	x	x
hsa-miR-183-5p	C02	UAUGGCACUGGUAGAAUUCACU	YP00206030	[127]	x	
hsa-miR-187-3p	C03	UCGUGUCUUGUGUUGCAGCCGG	YP00204018	[134]	x	x
hsa-miR-187-5p	C04	GGCUACAACACAGGACCCGGGC	YP00205920	[127]	x	
hsa-miR-192-5p	C05	CUGACCUAUGAAUUGACAGCC	YP00204099	[127]	x	
hsa-miR-193a-3p	C06	AACUGGCCUACAAAGUCCCAGU	YP00204591	[127]	x	
hsa-miR-193b-3p	C07	AACUGGCCCUCAAAGUCCCGCU	YP00204226	[134]	x	x
hsa-miR-194-3p	C09	CCAGUGGGGCUGCUGUUAUCUG	YP00204204	[127]	x	x
hsa-miR-194-5p	C08	UGUAACAGCAACUCCAUGUGGA	YP00204080	[127]	x	x
hsa-miR-196a-5p	C10	UAGGUAGUUUCAUGUUGUUGGG	YP00204386	[134]	x	x
hsa-miR-19b-3p	C11	UGUGCAAAUCCAUGCAAAACUGA	YP00204450	[134]	x	x
hsa-miR-200a-3p	D01	UAACACUGUCUGGUAACGAUGU	YP00204707	[127]	x	
hsa-miR-200a-5p	D02	CAUCUUACCGGACAGUGCUGGA	YP00206063	[127]	x	
hsa-miR-200b-3p	D03	UAAUACUGCCUGGUAAUGAUGA	YP00206071	[127]	x	x
hsa-miR-200b-5p	D04	CAUCUUACUGGGCAGCAUUGGA	YP00204144	[127]	x	
hsa-miR-200c-3p	D05	UAAUACUGCCGGGUAAUGAUGGA	YP00204482	[127]	x	x
hsa-miR-204-5p	D06	UUCCCUUUGUCAUCCUAUGCCU	YP00206072	[127]	x	x
hsa-miR-205-5p	D07	UCCUUCAUUCCACCGGAGUCUG	YP00204487	[127]	x	x
hsa-miR-210-3p	D08	CUGUGCGUGUGACAGCGGCUGA	YP00204333	[127]	x	x
hsa-miR-211-5p	D09	UUCCCUUUGUCAUCCUUCGCCU	YP00204009	[127]	x	x
hsa-miR-214-3p	D10	ACAGCAGGCACAGACAGGCAGU	YP00204510	[134]	x	x
hsa-miR-215-5p	D11	AUGACCUAUGAAUUGACAGAC	YP00204598	[127]	x	x
hsa-miR-21-5p	H05	UAGCUUAUCAGACUGAUGUUGA	YP00204230	[134]		
hsa-miR-224-5p	H08	UCAAGUCACUAGUGGUUCCGUUUAG	YP00204641	[213,214]	x	

hsa-miR-24-3p	H04	UGGCUCAGUUCAGCAGGAACAG	YP00204260	[213,214]	x	x
hsa-miR-25-3p	H10	CAUUGCACUUGUCUCGGUCUGA	YP00204361	[213,214]	x	x
hsa-miR-27b-3p	E01	UUCACAGUGGCUAAGUUCUGC	YP00205915	[134]	x	x
hsa-miR-29a-3p	E02	UAGCACCAUCUGAAAUCGGUUA	YP00204698	[134]	x	
hsa-miR-29b-3p	E03	UAGCACCAUUUGAAAUCAGUGUU	YP00204679	[134]	x	x
hsa-miR-29c-3p	E04	UAGCACCAUUUGAAAUCGGUUA	YP00204729	[134]	x	
hsa-miR-30a-3p	E06	CUUUCAGUCGGAUGUUUGCAGC	YP00204457	[127]	x	
hsa-miR-30a-5p	E05	UGUAAACAUCCUCGACUGGAAG	YP00205695	[127]	x	x
hsa-miR-30c-5p	E07	UGUAAACAUCCUACACUCUCAGC	YP00204783	[127]	x	x
hsa-miR-31-3p	E09	UGCUAUGCCAACAUAUUGCCAU	YP00204079	[127]	x	
hsa-miR-31-5p	E08	AGGCAAGAUGCUGGCAUAGCU	YP00204236	[127]	x	x
hsa-miR-320a	H07	AAAAGCUGGGUUGAGAGGGCGA	YP00206042	[213,214]	x	x
hsa-miR-331-3p	H11	GCCCCUGGGCCUAUCCUAGAA	YP00206046	[213]	x	
hsa-miR-340-3p	E10	UCCGUCUCAGUUACUUUAUAGC	YP00204250	[127]	x	
hsa-miR-342-3p	E11	UCUCACACAGAAAUCGCACCCGU	YP00205625	[127]	x	x
hsa-miR-345-5p	E12	GCUGACUCCUAGUCCAGGGCUC	YP00206006	[134]	x	x
hsa-miR-34a-5p	F01	UGGCAGUGUCUUAGCUGGUUGU	YP00204486	[134]	x	
hsa-miR-34b-3p	F02	CAAUCACUAACUCCACUGCCAU	YP00204005	[134]	x	
hsa-miR-361-5p	F03	UUAUCAGAAUCUCCAGGGGUAC	YP00206054	[127]	x	
hsa-miR-363-3p	F04	AAUUGCACGGUAUCCAUCUGUA	YP00204726	[127]	x	x
hsa-miR-372-3p	F05	AAAGUGCUGCGACAUUUGAGCGU	YP00204137	[134]	x	x
hsa-miR-373-3p	F06	GAAGUGCUUCGAUUUUGGGGUGU	YP00204604	[134]	x	
hsa-miR-375-3p	F07	UUUGUUCGUUCGGCUCGCGUGA	YP00204362	[127]	x	x
hsa-miR-382-5p	F08	GAAGUUGUUCGUGGUGGAUUCG	YP00204169	[134]	x	x
hsa-miR-423-5p	H09	UGAGGGGCAGAGAGCGAGACUUU	YP00205624	[213,214]	x	x
hsa-miR-485-5p	F09	AGAGGCUGGCCGUGAUGAAUUC	YP02112548	[127]	x	x
hsa-miR-506-3p	F10	UAAGGCACCCUUCUGAGUAGA	YP00204539	[127]	x	
hsa-miR-508-3p	F11	UGAUUGUAGCCUUUUGGAGUAGA	YP00204480	[127]	x	x
hsa-miR-509-3p	F12	UGAUUGGUACGUCUGUGGGUAG	YP00204458	[127]	x	x
hsa-miR-510-5p	G01	UACUCAGGAGAGUGGCAAUCAC	YP00204349	[127]	x	
hsa-miR-514a-3p	G02	AUUGACACUUCUGUGAGUAGA	YP00205931	[127]	x	x
hsa-miR-552-3p	G03	AACAGGUGACUGGUUAGACAA	YP00206032	[127]	x	x
hsa-miR-649	G04	AAACCUGUGUUGUUCAAGAGUC	YP00204146	[127]	x	x
hsa-miR-650	G05	AGGAGGCAGCGCUCUCAGGAC	YP00204233	[127]	x	
hsa-miR-661	G06	UGCCUGGGUCUCUGGCCUGCGCGU	YP00204657	[127]	x	x
hsa-miR-873-5p	G07	GCAGGAACUUGUGAGUCUCCU	YP00204175	[127]	x	
hsa-miR-92a-3p	G09	UAUUGCACUUGUCCCGGCCUGU	YP00204258	[134]	x	x
hsa-miR-92b-3p	G10	UAUUGCACUCGUCCCGGCCUCC	YP00204384	[134]	x	

hsa-miR-9-3p	G08	AUAAAGCUAGAUAACCGAAAGU	YP00204620	[134]	**x**	
hsa-miR-96-5p	G11	UUUGGCACUAGCACAUUUUUGCU	YP00204417	[127]	**x**	
hsa-miR-99a-5p	G12	AACCCGUAGAUCCGAUCUUGUG	YP00204521	[134]	**x**	**x**
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					

Evagreen-based Droplet Digital PCR and data analysis

We designed custom plates (96-well format) that included 89 different pre-spotted 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 [215] (miRNA list and plate set up in **Table 5**).

Droplet digital PCR workflow was performed using a EvaGreen-based miRNA quantification protocol that was recently developed [215-217] using a 20 µL volume of PCR mix containing 10 µL of 2X QX200 ddPCR EvaGreen Supermix (Bio-Rad), 6 µL of nuclease-free water and 4 µL of diluted cDNA. Each ddPCR mix was loaded into a disposable DG8 Cartridge (Bio-Rad) along with 70 µL of droplet generation oil for EvaGreen (Bio-Rad) in each oil wells. The cartridge was then covered with a DG8 Gasket (Bio-Rad) and placed inside the QX200 Droplet Generator (Bio-Rad). Generated droplets were then carefully transferred to a new semi-skirted 96-well plate (Bio-Rad) that was heat-sealed with a pierceable aluminum foil (Bio-Rad) into the PX1 PCR Plate Sealer (Bio-Rad) and placed in a T100 thermal cycler (Bio-Rad). Thermal cycling conditions were: 95 °C for 5', then 40 cycles of 95 °C for 30'' and 58 °C for 1' (ramping rate reduced to 2%), and three final steps at 4 °C for 5', 90 °C for 5' and a 4 °C infinite hold. Droplet selection was performed individually for each well using QuantaSoft software v 1.7 (Bio-Rad). Specifically, from the 2D amplitude plot we selected the positive droplets in each well using the lasso tool. Using the events tab, the number of positive and total generated droplets were observed. The general performance of EvaGreen ddPCR allowed to obtain a total number of 18,000–21,000 droplets per well.

Final miRNA amounts (copies/µl) were obtained and normalized on 50th percentile expression using GX v.14.9.1 software (Agilent Technologies). Candidate reference RNAs were not used as normalizer in this analysis due to the higher variability compared to the median expression.

Tissue-of-origin prediction

Primary tumors (N=96) were used as reference set as previously described [127]. Primary tumors deriving from the same patient were averaged prior of normalization on 50th percentile expression. Nearest shrunken centroids (NSC) algorithm [218] using the Prediction Analysis of Microarray for R (PAMR) tool [218] and the Least Absolute Shrinkage and Selection Operator (LASSO) model [219] were used to build up the classifiers and select the discriminant miRNAs for primary site prediction [212]. In detail, PAM applies a shrinkage nearest neighborhood centroid approach in the space of the samples. In the reference 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 towards zero by a quantity called threshold that is set by the user to minimize the error rate according to the results of cross-validation technique. If the shrinkage process reduces to zero the centroid of a gene across all the classes, the specific miRNA is not selected for the prediction. Then in the prediction step the distance between miRNA 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.

On the other hand, 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 towards 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.

We set PAM threshold to 0 leading to a classifier based on 87 miRNAs while the LASSO threshold was set to 0.019 leading to a classifier based on 53 miRNAs (miRNAs are listed in **Table 5**). Then, these classifiers were used to predict known and unknown/uncertain metastases tissue-of-origin. Both models assigned to every metastatic tumor a probability to be originated from each primary tumor class. The variable gender was also taken into account to exclude not compatible molecular

predictions (TGSC/PRAD in females and OV/UCEC in males). Prediction outcome was compared with the indications of a possible primary site suggested by standard diagnostic workup and clinico-pathological assessment. Bootstrap approach (with N=100) was used to assess the performance (error rate) of the models in the training set.

TCGA data download, filtering and prediction

Samples from 8 out of 17 tumor types included in this study were present in the TCGA data (BLCA, CHOL, BRCA, LIHC, LUAD, LUSC, OV, PAAD) along with their matched normal tissues. For these 8 tumor types and their normal counterpart we selected our 89 miRNAs using firebrowseR R package. Of note, even though we were aware that BRCA is wider than LBC class, we decided to include in this analysis anyway. Then, on the whole matrix we applied a two-step filtering process to select samples and miRNAs and exclude missing values. First, we selected samples with expression values detectable in at least 80% of the miRNA set, and secondly, we select miRNAs without missing values in the selected sample set. We end up with 835 patients and 48 mRNAs. LIHC tumor samples were excluded from this analysis due to the low quality of these data. The same procedure has been applied for normal tissues obtaining 1533 samples and 47 miRNAs. TCGA data from both normal and tumor samples were used to perform primary-site prediction with PAM and LASSO.

Cluster analysis

We performed cluster analysis using the normalized expression (50^{th} percentile) of the 89 miRNAs in a) individual patients of the reference set of primary tumors and b) averaged levels within each tumor class of the reference set. Both hierarchical cluster analyses were performed using GeneSpring GX v.14.9.1 software (Agilent Technologies) using complete-linkage rule and Manhattan correlation distance.

Survival analysis

Univariate survival analysis was performed using Kaplan-Meier curves and the log rank test, as implemented in survMisc R package. Overall survival (OS) was measured considering the time lagging between diagnosis and death for any cause or the last follow-up. For each miRNA the optimal cut-off was estimated as the threshold on the ROC curves that maximize the sum of specificity and sensitivity in

predicting CUP patients. Results were reported as *p* value, hazard ratio (HR) and 95% confidence intervals (CI). A *p* value $\leqslant 0.05$ was considered significant.

CTC isolation with CELLSEARCH and DEPArray NxT

CB055 and CB090 blood samples were collected at diagnosis. CELLSEARCH® system (Menarini Silicon Biosystems) was used to enrich and enumerate CTCs, followed by single cell isolation with DEPArray™ NxT system (Menarini Silicon Biosystems). Briefly, 7.5 ml of blood were loaded in CELLSEARCH Autoprep® for EpCAM-based immunomagnetic capture of CTCs, followed by immunofluorescent staining (CK-FITC, PD-L1-PE, CD45-APC and DAPI using the CELLSEARCH CXC kit in one patient, and CK-PE, CD45-APC and DAPI with the CELLSEARCH CTC kit in the other patient). After CTCs enumeration with the CellTrack Analyzer II, the sample was loaded into a DEParray Cartridge and single CTCs and leukocytes were isolated with the DEPArray NxT system (Menarini Silicon Biosystems) from each sample. A variable number of CTCs and leukocytes, used as control, were isolated as single cells and then subjected to whole genome amplification employing the Ampli1 WGA Kit (Menarini Silicon Biosystems).

CTC isolation with Parsortix system

For CB090, tumor cells were parallelly enriched from peripheral blood also using Parsortix system (ANGLE plc). Blood cells were forced to pass through a 6.5 µm separator cassette that can capture CTCs based on their size and deformability properties. Then, to confirm the CTC nature of retained cells, we stained CTCs using a combination of anti-human EpCAM-FITC, anti-human CD45-APC (BD 561864/555485) and DAPI for cells nuclei staining. Images were acquired using a Leica DMI6000 B inverted fluorescence microscope (Leica Microsystems).

SureSelect analysis of tumor genetic alterations

A 1.2 Mb custom SureSelect capture bait library (Agilent Technologies) was designed to target 92 genes (listed in **Table 6**), selected to be the most frequently mutated genes in CUPs, as reported in MSK Impact study CUP cohort [152], including druggable or potentially actionable alterations.

The panel was designed using Agilent SureDesign web application (https://earray.chem.agilent.com/suredesign/, v. 7.0, Agilent Technologies) to cover coding exons and UTRs, with 25 flanking bases from 3' and 5' ends. This panel

comprehends up to 51,466 probes included in five probe groups, with 2x tiling density, balanced boosting and 86% of probes with the most stringent masking. Libraries were prepared using the custom gene panel following SureSelectXT HS / SureSelectXT Low Input Target Enrichment with Pre-Capture Pooling protocol (G9702-90005, v. A0, June 2019, Agilent Technologies). An input of 50 ng of gDNA (normal and tumor) underwent to enzymatic fragmentation according to SureSelect Enzymatic Fragmentation protocol (G9702-90050, Revision B0, January 2020, Agilent Technologies). Libraries from ccfDNAs of CB055 (A, C) and CB090 were prepared starting from 25 ng of DNA bypassing the fragmentation step, as describe in a published application note [220].

Table 6. List of genes and regions included in SureSelect custom panel [221].

Gene	Covered region
ACD	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ALK	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
APC	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ARID1A	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ARID1B	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ASXL2	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ATM	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ATR	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ATRX	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
BAP1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
BCOR	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
BLM	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
BRAF	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
CARD11	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
CCNE1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
CDK4	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
CDKN2A	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
CDKN2B	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
CIC	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
CREBBP	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
CTNNB1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
DOT1L	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
EGFR	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
EP300	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
EPHA3	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
EPHA5	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
EPHA7	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ERBB2	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ERBB3	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ERBB4	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')

FAT1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
FBXW7	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
FGFR2	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
FLT4	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
GRIN2A	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
HGF	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
HIST1H1C	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
HNF1A	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
INPP4B	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
JAK3	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
KDM5A	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
KDR	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
KEAP1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
KIT	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
KMT2C	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
KMT2D	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
KRAS	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
MC1R	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
MCL1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
MDC1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
MED12	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
MET	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
MGA	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
MITF	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
MTOR	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
MYC	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
NF1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
NF2	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
NOTCH1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
NOTCH2	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
NOTCH3	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
NOTCH4	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
NRAS	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
NTRK1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
NTRK3	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
PALB2	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
PBRM1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
PIK3C2G	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
PIK3CA	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
PIK3CG	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
PLCG2	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
POLD1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
POT1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
PPP6C	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
PTPRD	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
PTPRT	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
RB1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
RBM10	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')

RET	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
RHOA	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ROS1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
RPTOR	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
SETD2	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
SMAD4	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
SMARCA4	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
SPEN	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
STAG2	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
STK11	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
TERT	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
TGFBR1	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
TP53	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
ZFHX3	coding exons + 5' and 3' UTR (25 bp of flanking bases at 3' and 5')
TMEM38B	rs10739221
TYR	rs1126809
CCND1	rs12422135, rs2290419, rs498136
FTO	rs12596638
OCA2	rs145720174
AGR3	rs1636744
SLC45A2	rs16891982
PARP1	rs2249844, rs1858550
GC	rs2282679
CASP8	rs2349073
OBFC1	rs2995264
ARNT	rs3768013
MX2	rs443099
RMND2	rs6750047
CDKALI	rs6914598
TYRP1	rs72706189
ASIP	rs74325991, rs2377956
IRF4	rs9405705

Libraries were sequenced on NextSeq 500 (Illumina) platform using a High Output 2x75 bp flow cell. Variant calling and paired analyses (tumor vs normal) were performed using SureCall software (v. 4.2), applying a filter for gDNA at 5%, and at 1% for ccfDNAs. Somatic alterations were filtered to only keep exonic nonsynonymous single nucleotide variants (SNVs), insertions, deletions, multiple nucleotide variants, and long deletions not detected in the normal sample.

Alterations with population allele frequency (Genome Aggregation Database, GnomAD [222]) higher than 0.5% in Non-Finnish European (NFE) population were also filtered out. All variants reported had a coverage higher than 100. Target sequencing data (92-gene custom panel) were used to assess copy number variants using panelcn.MOPS pipeline [223]. Variants triage and clinical

classification was obtained with Alissa Interpret software (Agilent Technologies). Variants were also annotated using ANNOVAR [224]. Bioinformatic pathogenicity prediction, reported in **Table 13** and **Table 15,** of the identified variants was performed consulting the prediction score/outcome of SIFT (Sort Intolerated From Tolerated [225]), Polyphen2 HVAR (Polymorphism Phenotyping v2 [226]), LRT (Likelihood Ratio Test [227]), Mutation Taster [228] , Mutation Assessor [229], FATHMM (Functional Analysis Through Hidden Markov Model [230]), CADD (Combined Annotation Dependant Depletion [231]) and VEST (Variant Effect Scoring Tool [232]). Genes with evidence for actionability and the FDA approved drugs reported in **Table 13** and **Table 15** were obtained consulting OncoKB [157].
Single–cell genomic DNA from DEPArray NxT sorted cells, amplified using Ampli1™ WGA Kit, was processed to obtain Illumina-compatible libraries using the Ampli1™ OncoSeek Panel and the Ampli1™ LowPass kits. On the other hand, tumor gDNA was directly processed with DEPArray™ LibPrep and DEPArray™ OncoSeek Panel kits. After sequencing on MiSeq platform, raw data were analyzed using assay-specific applications on the MSBiosuite platform (Menarini Silicon Biosystems).
FFPE tumor tissue of CB055 was also eligible for genetic testing using FoundationOne CDx (F1CDx) assay. F1CDx is performed exclusively as a laboratory service using DNA extracted from formalin-fixed, paraffin-embedded (FFPE) tumor samples. Library were prepared with a hybrid capture-based target enrichment approach and whole-genome shotgun library construction to detect substitutions, indels, copy number alterations, and selected rearrangements in a total of 324 genes. Sequencing was performed using the Illumina HiSeq 4000 platform, obtaining a >500X median coverage. F1CDx assay simultaneously assessed tumor mutation burden (TMB) and microsatellite instability (MSI) status of the analyzed specimen (Further technical information is available at https://www.foundationmedicine.com/genomic-testing/foundation-one-cdx).

Probe-based Droplet Digital PCR validation

Droplet Digital PCR technology (Bio-Rad) along with Mutational and Copy Number Determination assays (Bio-Rad) were used to validate specific alterations in both genomic and circulating cell-free DNAs of CB055. A Copy Number Variation assay was used to assess the amplification of FGFR2 gene (dHsaCB2500320, Bio-Rad) using RPP30 as reference gene (dHsaCP2500350, Bio-Rad). A digestion with Hae III restriction enzyme was performed during ddPCR, according to manufacturer's

instruction. Mutational assay for wild type and mutated ARID1A (p.R1276*) was performed with a custom IDT assay (02516372). For this assay, DNA digestion with Hind III restriction enzyme was performed only for gDNA. Probe-based Droplet digital PCR experiment was performed according to manufacturer's instruction and thermal cycling conditions were: 95 °C for 10 minutes, 94°C for 30 seconds and 58°C for 1 minute for 40 cycles, 98 °C for 10 minutes and an infinite hold at 4 ° C (ramping rate reduced to 2%). Droplet selection was performed individually for each well using QuantaSoft software v 1.7 (Bio-Rad) obtaining the final copy number variation or the wild type/mutated allele fractions.

Results

Part 1: Molecular diagnosis and prognosis of CUP using a microRNA-based droplet digital PCR assay

Multi-miRNA testing on archive samples with droplet digital PCR

MiRNAs are extremely stable in formalin-fixed paraffin-embedded tissues, which are the most common source of tumor material available in the clinical setting. Therefore, we designed an on-demand tool able to test the absolute levels of 89 miRNAs in a 2-days timeframe, confirming its feasibility during standard diagnostic workup.

The absolute quantification of miRNAs was performed using EvaGreen-based Droplet Digital PCR technology according to a recently published protocol [215]. The assay achieved good quality results for all the FFPE samples analyzed, with no sample excluded for technical reasons.

RNA was extracted from 2-5 FFPE tissue slices by macrodissecting the tumor area that experienced pathologists carefully selected. We were able to test all miRNAs in a single experiment using only 10 ng of RNA.

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

With the aim of establishing a reference set for cancer of unknown origin molecular profiling, we tested 96 primary tumors with our multi-miRNA assay, comprising 16 different tumor types and up to 19 histological classes. Primary tumor classes were selected according to the most common CUP's sites of origin identified at autopsy [233]. We obtained the expression matrix of the primary tumor dataset, constituted by tumors belonging to 19 different classes: LUAD, LUSC, PAAD, LIHC, CHOL, KIRC, KIRP, STAD, CRC, TGSC, OV, UCEC, BLCA, LBC, TNBC, PRAD, SKCM, GI-NET and HNSC. An overview of the primary tumor samples for each histological subtype included in this study is reported in **Table 7**.

Table 7. Summary of samples and patients enrolled in the study [212].

		Primaries *(reference set)*		**Metastases** *(test set)*		**CUPs**	
Characteristics		**n**	***%***	**n**	***%***	**n**	***%***
Patients	n = 150	94		10		46	

Prospective						10	*22*
Retrospective		94		10		36	*78*
Samples	n = 159	96		10		53	
Sex							
Male	78	48	*50*	3	*30*	26	*49*
Female	64	32	*33*	5	*50*	27	*51*
ND	17	16	*17*	2	*20*	0	*0*
Age, years							
Median		66		71		67	
Range		44-85		60-86		42-87	
ND		62		4		0	
Primary tumor classes:							
BLCA		4					
CHOL		6					
CRC		7		1			
GI-NET		5					
HNSC		6		1			
KIRC		5		1			
KIRP		3					
LBC		5		1			
LIHC		6					
LUAD		6		1			
LUSC		3					
OV		6					
PAAD		5		1			
PRAD		5		1			
SKCM		7		1			
STAD		5		1			
TGSC		4					
TNBC		3					
UCEC		5		1			
Metastatic sites:							
BONE						2	
BONE MARROW						1	
BRAIN				1		2	
BREAST						3	
CEREBELLUM						1	
COLON				1		1	
DERMIS						1	
DUODENO						1	
KIDNEY						1	
LIVER				4		12	
LUNG				1		2	
LYMPH NODE						14	
MUSCLE						1	

ND		1
PERICARDIUM	1	
PLEURA		5
PROSTATE		2
SKIN	1	
SOFT TISSUES		2
STOMACH	1	
THYROID		1

BLCA: transitional cell carcinoma of bladder; CHOL: cholangiocarcinoma; CRC: colorectal adenocarcinoma; GI-NET: gastrointestinal neuroendocrine carcinoma; HNSC: head and neck squamous cell carcinoma; KIRC: kidney renal clear cell carcinoma; KIRP: kidney renal papillary cell carcinoma; LBC: luminal non-special type and lobular breast carcinoma; LIHC: hepatocellular carcinoma; LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma; OV: ovarian serous carcinoma; PAAD: pancreas exocrine adenocarcinoma; ; PRAD: prostate adenocarcinoma; SKCM: melanoma of skin; STAD: gastric adenocarcinoma; TGSC: germ cell seminomatous carcinoma; TNBC: triple negative breast cancer; UCEC: endometrial adenocarcinoma; ND: not defined.

Analysis of miRNA expression patterns

Cluster analysis was performed using the normalized expression of the 89-miRNA signature to evaluate the average miRNA levels in each tumor class included in the reference set (**Figure 7**) and their absolute levels in individual patients belonging to the reference set (**Figure 8**).

As expected, each tumor class shows a distinct pattern of miRNA expression. However, we noted that STAD and CRC had similar miRNA expression profiles and were partially intermixed with other gastrointestinal tumors (PAAD and GI-NET), as reported elsewhere [127,133,134]. For this reason, we decided to consider them as a single class (STAD-CRC) for molecular prediction. Similarly, our analysis reported overlapping miRNA expression patterns in kidney renal clear cell (KIRC) and papillary cell carcinomas (KIRP), so we decided to combine them in the tumor class KICA. Similar but yet distinct miRNA patterns were observed in tumors in female reproductive-system organs (OV and UCEC), as reported previously [127,133,134]. Unexpectedly, lung cancers (both LUAD and LUSC) were found to share a portion of their signatures with TNBC but not with other breast cancer subtypes (ER+, PR+, HER2+ tumors). In fact, TNBC present a divergent pattern of miRNA expression when compared to other breast cancers and shows a surprising similarity with HNSC instead. A common etiology associated to human papilloma virus (HPV) infection has been reported in both these tumor types [234-236].

Overall, this signature confirmed its potential in discriminating among up to 17 different tumor classes.

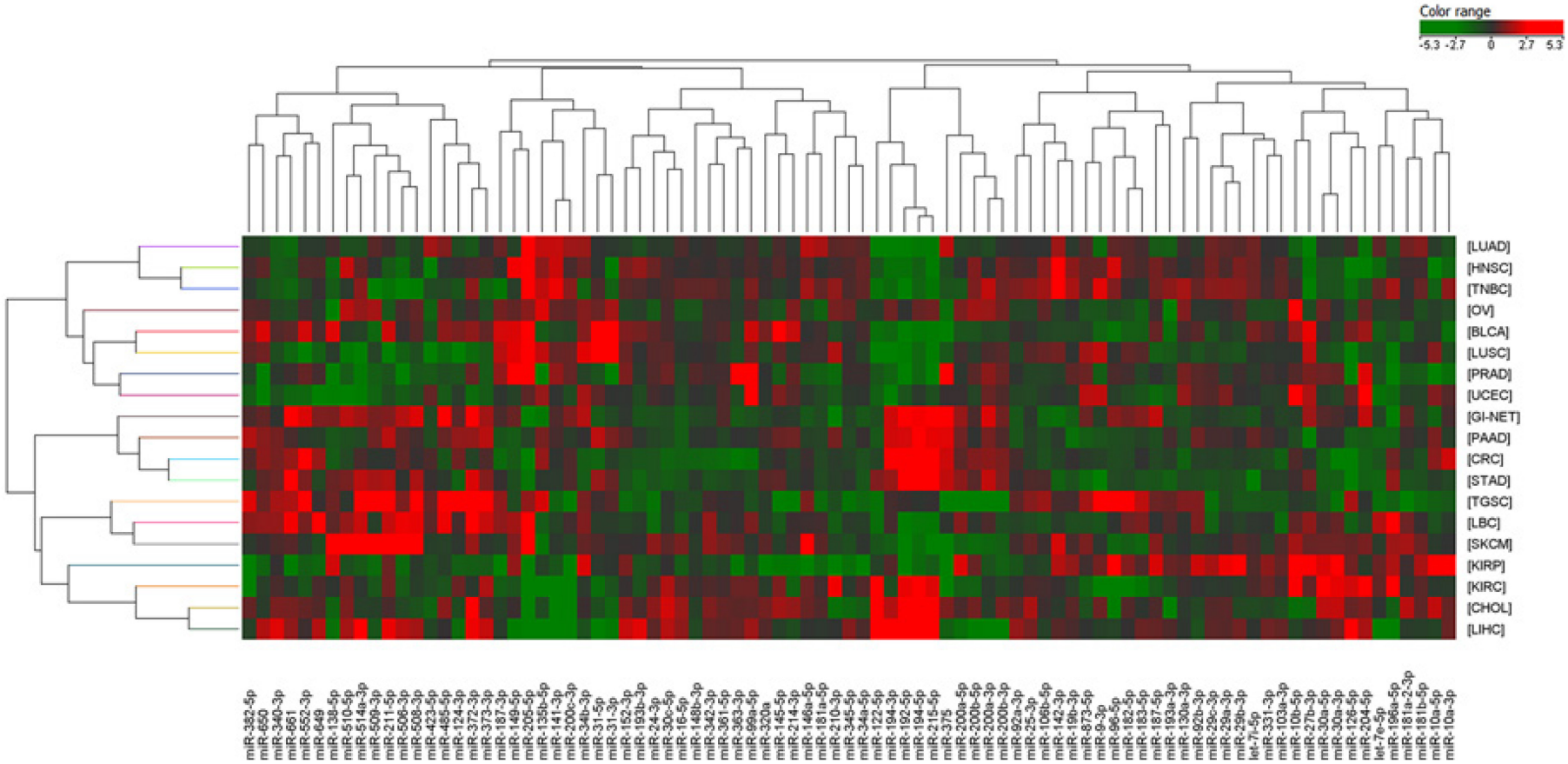

Figure 7. Cluster analysis of averaged expression per class of the 89 miRNAs in the reference set. Heatmap representing the expression of the 89-miRNA signature in nineteen different tumor types. Normalized miRNA levels, averaged for each tumor class were used for clustering analysis. Green indicates low expression, red indicates high expression.

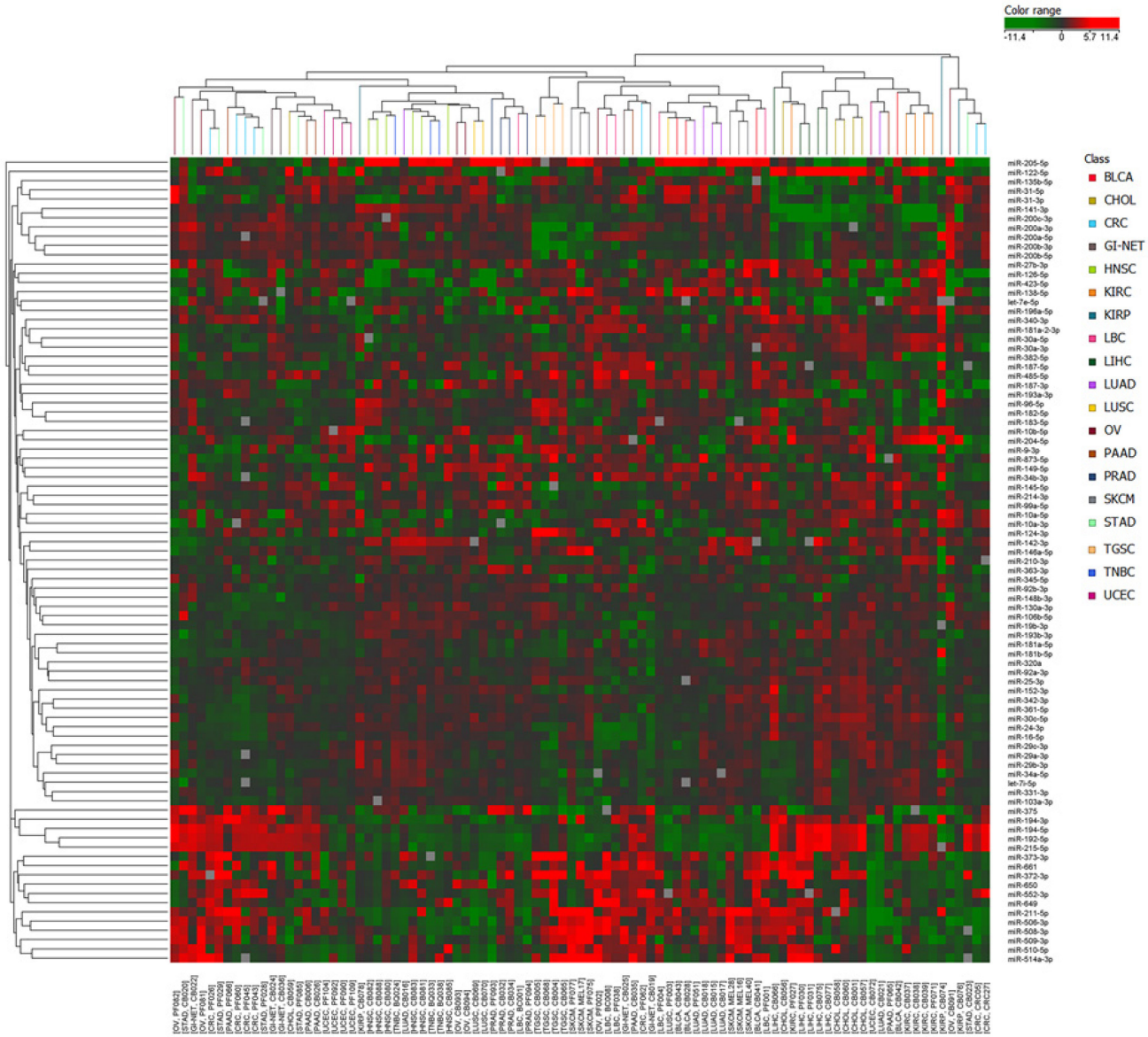

Figure 8.

Cluster analysis on individual patients of the training set.

Heatmap representing the expression of 89 microRNAs in 94 samples of the training set belonging to nineteen different classes of primary tumors. Normalized miRNA levels for each sample were used for clustering analysis. Green indicates low expression, red indicates high expression.

CUP predictive model generation

Final tissue-of-origin prediction was performed using 87 out of 89 miRNAs of our panel. In particular, miR-122-5p was excluded from the prediction analysis due to its strong signal generated by the liver microenvironment in metastatic samples (as displayed in **Figure 9**), while miR-21-5p was omitted due to its lack of specificity with both classifiers (it is widely expressed in solid tumors).

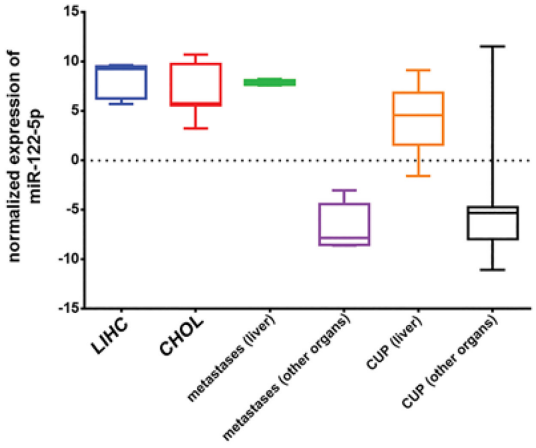

Figure 9. Plot of miR-122-5p expression in primary and metastatic tumors.
We compared normalized expression of miR-122-5p in LIHC, CHOL and metastases of known/unknown origin in which the sample biopsy was obtained from liver or other organs. High levels of miR-122-5p were reported in LIHC and CHOL, as expected. Since miR-122-5p is the most abundant liver-derived miRNA, in liver cells and in the tumor microenvironment, its high expression was also observed in metastatic tumors of known/unknown origin whose biopsy was obtained from the liver tissue, if compared to those derived from other sites.

We applied two predictive models developed by *Tibshirani et al.* to our reference set of primary tumors: specifically, the Nearest Shrunken Centroids (NSC) using PAMR [218] and the Least Absolute Shrinkage and Selection Operator (LASSO) models [219]. We assessed their performance on the training set using a bootstrap approach. Error rates for each tumor class for both models are reported in **Table 8**. Of note, the overall error rate for both PAMR and LASSO was 33%. However, 11 among the 17 tumor classes included in the reference set (LIHC, LUSC, LBC, KICA, GI-NET, TGSC, STAD-CRC, SKCM, LUAD, UCEC, PRAD) reported error rates much lower with both models (17% for PAMR and 22% for LASSO). Notably, we observed that PAMR seem to be much more precise in identifying LBC and LUSC compared to LASSO; on the contrary, LASSO showed a higher accuracy in the classification of UCEC, LUAD and SKCM. Both models had problems in predicting correctly BLCA, PAAD, TNBC, HNSC, OV and CHOL classes; this might be caused

by a reduced specificity of the 87-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). These results highlight that these two models behave similarly on some classes and complementarily in some others, therefore we decided to take advantage of both classifiers and combine their molecular prediction.

Table 8. Error rates of the two models for each tumor class [212].

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
overall error rate	0.33	0.33

In addition, a small set of metastases of known origin (N=10) was assessed for molecular prediction (test set in **Table 7**). Considering the two top predicted classes we obtained an accuracy of 70% for both predictive models as reported in **Table 9**.

Table 9. Error rates of PAMR and LASSO in metastases of known origin [212].

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

In particular, PAMR failed in the identification of the right primary site when dealing with UCEC, PRAD and KIRC, while LASSO misclassified UCEC, PRAD and SKCM. In addition, we evaluated the performance of both classifiers on 8 tumor classes included in TCGA database (BLCA, BRCA, CHOL, LIHC, LUAD, LUSC, OV and PAAD), testing both tumor and matched normal samples. Despite available TCGA miRNA expression data with adequate quality signal were available for only 48 miRNAs in tumor samples (7 tumor classes) and 47 miRNAs in normal samples (8 tumor classes), PAMR and LASSO predictions showed an overall median positive prediction rate (PPR) higher than 80% for both tumor and normal samples (**Table 10**).

Table 10. Positive prediction rates of LASSO and PAMR in tumor and normal tissue samples based on TCGA data [212].

	PPR [a]			
tumor classes	**LASSO tumor**	**LASSO normal tissue**	**PAMR tumor**	**PAMR normal tissue**
BLCA	100%	97%	94%	96%
CHOL	100%	100%	100%	100%
BRCA	67%	97%	100%	94%
LIHC	-	100%	-	92%
LUAD	85%	79%	100%	74%
LUSC	50%	93%	27%	63%
OV	100%	100%	100%	86%
PAAD	100%	100%	100%	60%
total	86%	96%	89%	83%

[a] *positive prediction rate*

CUP primary site prediction

Finally, both models were used to infer the tissue-of-origin of a remarkably large collection of 53 cancers of unknown/uncertain origin (CUPs) (**Table 7**). The prediction outcome is displayed in **Figure 10** in which the top two primary sites predicted by both models for each CUP sample are reported. Using PAMR, the molecular prediction of 81% of CUPs (N=43) reached a probability higher than 60%, and 55.8% of them even higher than 90%. Using LASSO, the molecular prediction of 47.2% of CUPs (N=25) predicted the first primary site with a probability higher than 60%, and 13% (N=7) higher than 90%.

The most probable tissues-of-origin, reported in **Table 11,** were prioritized using the following criteria: the tissue-of-origin was a) predicted by at least one predictive

model (LASSO or PAMR) with a probability higher than 80%, b) 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 (including the first and second predictions). Following the prioritization, a probable tissue-of-origin was assigned to each CUP. Few cases had more than one tissue-of-origin. Of note, PAMR and LASSO predictions had high agreement: specifically, the same primary sites (according to the above-mentioned criteria) were predicted by both models in 94% of cases.

We also assessed the consistency of this molecular prediction with the clinico-pathological 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 formulated. Moreover, in 3 patients of our CUP cohort (CB071, CB054 and CB098/100/101/102), the primary site was later identified, allowing us to observe a full agreement of the molecular predictions with the final diagnoses.

We recognized a subgroup of CUP samples (N=5) in which we found very challenging to indicate a tissue-of-origin using both models, with molecular predictions with a probability ≤ 40%. This result could be explained by an exceptionally undifferentiated phenotype of these tumors or their primary tumors were not included in the reference set. Considering the final predicted sites reported in **Table 11**, the most common primary tumors were STAD-CRC (19%), LBC (15%), PAAD (15%), LUAD (13%), CHOL (11%), LUSC (5%), TNBC (8%) and HNSC (5%) and others at lower rates. Of note, no CUP was predicted as TGSC or PRAD.

Interestingly, we obtained spatially distinct synchronous and metachronous metastases (N=2-4) from 5 CUP patients included in our cohort. All metastases were tested with our molecular assay to evaluate the consistency of our prediction among different metastatic sites. Peculiar is the case of a patient (#B) with an initial diagnosis of CUP (later attributed to a breast origin) from which we obtained up to four samples (CB098, CB100, CB101, CB102). In particular, CB098 and CB100 were obtained from two different lymph nodes resected in 2010, while CB101 and CB102 derived from an invasive ductal breast cancer identified two years later, which was recognized as the primary site. Both PAMR and LASSO predictions agreed to point out an LBC or TNBC origin (**Table 11**). However, CB100 was predicted as LUAD (first) or TNBC (second) by PAMR classifier (**Figure 10**), probably due to the lower tumor cell fraction in this sample and the well-known similarity in miRNA expression between breast and lung cancers [127].

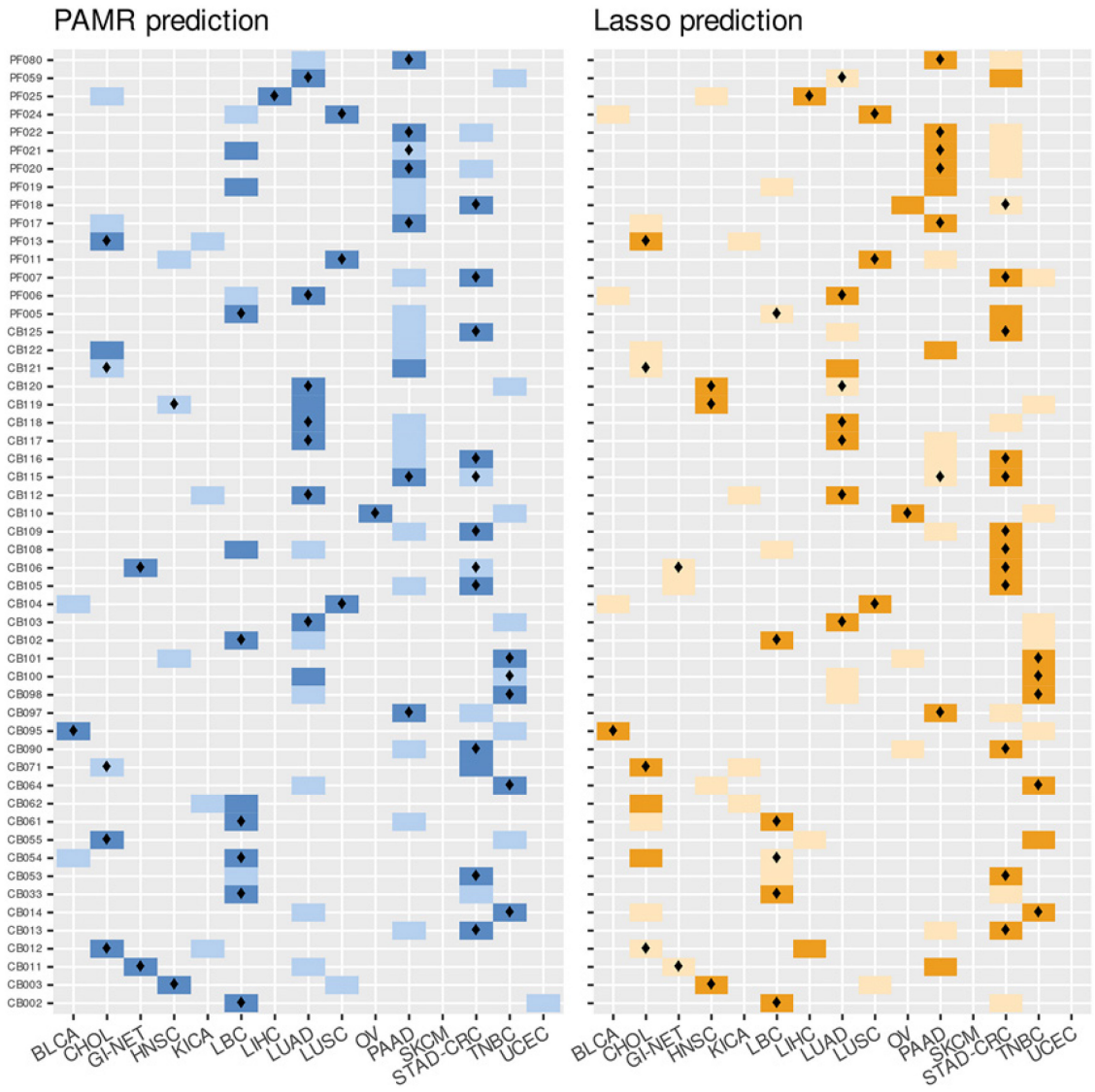

Figure 10 – Prediction outcome of CUPs using PAMR and LASSO classifiers. For each of the 53 CUP samples (on the y-axis) the two top predicted primary tumors (x axis) are highlighted. PAMR first and second molecular predictions are reported with dark and light-blue squares, respectively. LASSO first and second molecular predictions are reported in dark- and light-orange, respectively. A diamond in the square indicates those predictions that are compatible with pathological and/or clinical information.

Molecular predictions for the multiple metastases of the other patients (#E, #F, #Q and #R) reported concordant results for both models, consistent with clinico-pathological hypotheses. Specifically, for #E (CB105 and CB106) and #F (CB108 and CB109) both models agreed to predict a gastrointestinal origin (STAD-CRC), which were also the first pathological hypotheses. Of note, CB108 PAMR prediction pointed out to LBC, however, being derived from the bone it is probable that the sample had a compromised integrity. Molecular prediction for #R (CB121 and

CB122) pointed out to a biliopancreatic origin, while for #Q (CB119 and CB120) the two metastatic samples were predicted to have the same origin, which was, in this case, lung or head and neck.

Table 11. Prediction outcome in cancer of unknown primary site [212].

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-, synaptophysin-, CD56-, HER-2-, MSI-	gastrointestinal	ND		CHOL
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-, synaptophysin-, TTF1-	gastrointestinal	ND		STAD-CRC[†]
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 90%, Mammaglobin-, TTF1-, ER-, PR-, AR-, HER2-, BRAF V600E-, EBV-(ISH)	breast	breast	TNBC	LUAD / TNBC[†]
CB101	F	B03	42	80	retrospective	breast	adenocarcinoma	**POS**	**POS**	**POS**	CD10-, CEA+, GATA3-, HER2-, Ki67 70%, Mammaglobin-, CatepsinK-, 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-, CatepsinK-, 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-, CD10-, CD117-, CD56-, Ki67 50%, EBV- (ISH)	lung	ND		LUAD
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-, PDPN-, TTF1-	ND	ND		LUAD[†]

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 -, HER2- , Ki67 85%, EBV- (ISH)	breast	ND		LUAD
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

‡*The most probable primary sites according to the above-mentioned criteria;* † *suboptimal samples deriving from bone metastases or with a tumor cellularity ≤40%; ND: not defined.*

Association of microRNAs with CUP patients' overall survival

We evaluated the potential prognostic role of the 87 miRNAs included in our panel, in 34 CUP patients in which overall survival (OS) information was available. Survival analysis reported a significant association of miRNA expression with survival probability for 13 miRNAs (**Table 12** and **Figure 11**). In particular, this association was negative for 5 miRNAs (HR > 1) and positive for 8 miRNAs (HR < 1).

Table 12. Association of miRNA expression with overall survival (significant miRNAs) [212].

miRNA	**HR** [a]	**lower 95%**	**upper 95%**	***p*-value** [a]
miR-124-3p	0.11	0.03	0.36	**0**
miR-9-3p	0.29	0.12	0.71	**0.01**
miR-149-5p	0.32	0.13	0.78	**0.01**
miR-372-3p	0.33	0.12	0.89	**0.03**
miR-485-5p	0.37	0.16	0.9	**0.03**
miR-375	9.6	1.3	73	**0.03**
miR-25-3p	0.26	0.08	0.87	**0.03**
miR-27b-3p	2.6	1.1	6.1	**0.03**
miR-181a-2-3p	0.38	0.15	0.93	**0.03**
miR-10b-5p	0.35	0.13	0.93	**0.04**
miR-96-5p	2.5	1	6.2	**0.04**
miR-423-5p	3.5	1	12	**0.04**
miR-214-3p	2.6	1	6.6	**0.05**

[a] *hazard ratio (HR) with 95% confidence interval and p-value (p) for OS.*

Specifically, miRNAs whose higher expression was associated with a worse prognosis were miR-375 (p=0.03), miR-27b-3p (p=0.03), miR-96-5p (p= 0.04), miR-423-5p (p=0.04) and miR-214-3p (p= 0.05). On the contrary, miRNAs that were positively associated with a prolonged survival were miR-124-3p (p= 0.0002), miR-9-3p (p= 0.01), miR-149-5p (p= 0.01), miR-372-3p (p= 0.03), miR-485-5p (p= 0.03), miR-25-3p (p= 0.03), miR-181a-2-3p (p= 0.03) and miR-10b-5p (p= 0.04).

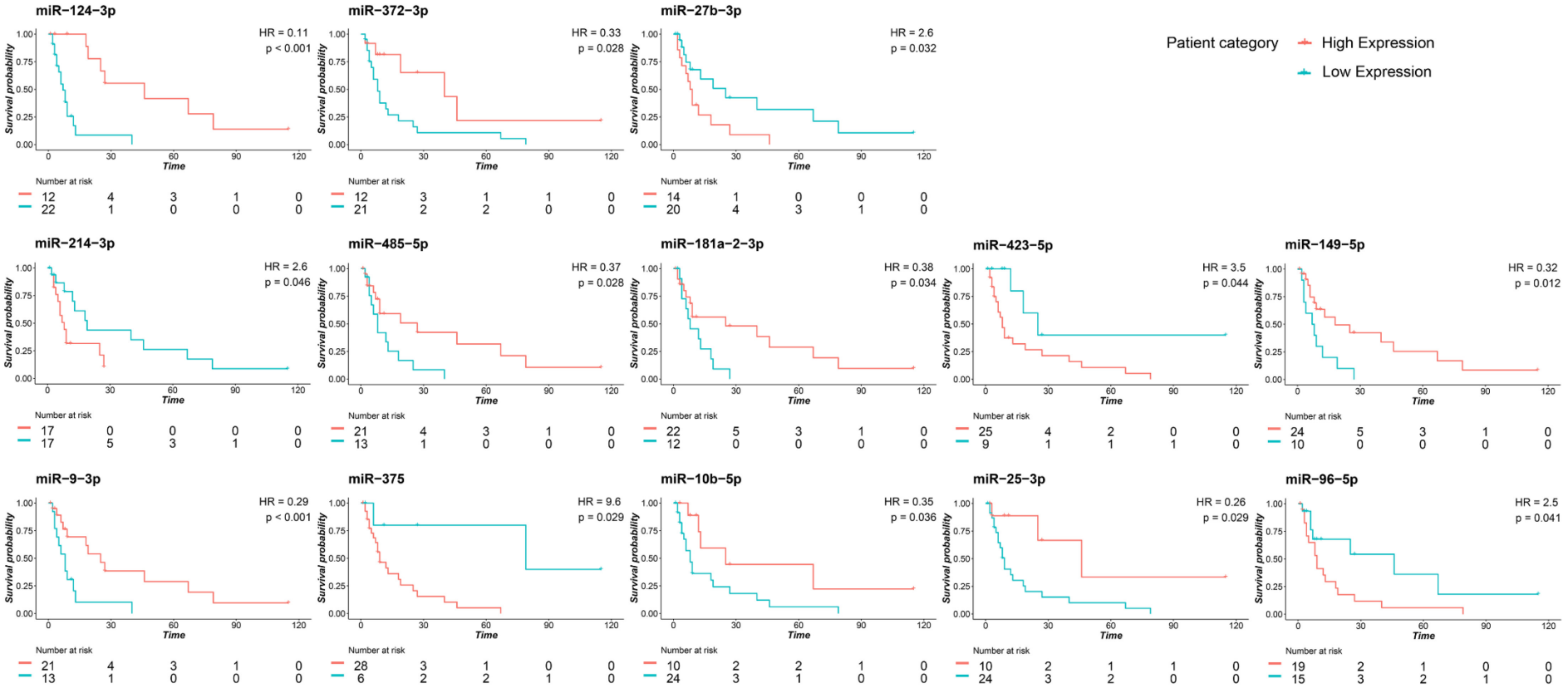

Figure 11 - **Kaplan–Meier OS curves based on the expression of 13 miRNAs in CUP patients.** Survival plots showing significantly different OS curves in high- (red lines) and low- (blue lines) miRNA expressing CUPs. The long-rank test was used to compare the survival distributions. The threshold for each miRNA was established based on the best performing value at ROC analysis. For 5 miRNAs, a higher expression was associated with a shorter survival, and for 8 miRNAs a higher expression was associated with prolonged survival of CUP patients. The x-axis represents the months from the diagnosis.

Part 2: Potential utility of liquid biopsy in the management of two prospective CUP patients

Case presentation: CB055

We then focused our study on two prospective CUP cases included in our cohort, CB055 and CB090, to evaluate the value of the application of liquid biopsy in the management of these patients [221].

Specifically, CB055 was a 49-years-old woman with a history of primary biliary cholangitis (PBC) that was admitted at Bologna University Hospital in June 2018 due to enlarged left cervical lymph nodes associated with severe asthenia, pruritus and fever. Contrast-enhanced computed tomography (CT) and F-Fluorodeoxyglucose positron emission tomography/computed tomography (18-F-FDG PET/CT) scan showed multiple pathologic abdominal and left supraclavicular lymph nodes. Two lymph nodes were resected from CB055 by abdominal surgery and pathological examination highlighted the presence of nodal metastasis from adenocarcinoma. Immunohistochemistry (IHC) staining reported the positivity for keratin 7 (K7), keratin 20 (K20) and CDX2, and negativity for neuroendocrine markers (chromogranin, synaptophysin, CD56), which pointed out to a likely lymph node metastasis from adenocarcinoma of the gastrointestinal tract.

However, esophagogastroduodenoscopy (EGDS) and pancoloscopy (PC) were not able to detect any sign of gastrointestinal neoplastic lesion. Molecular analyses performed in the clinical setting were all negative: specifically, the evaluation of microsatellite instability (MSI), HER2, ALK and PD-L1 status (by IHC), *KRAS*, *NRAS*, *BRAF* mutational status and *MGMT* methylation status (by pyrosequencing), and *ROS1* and *MET* (by FISH). Moreover, the only positive serum oncomarkers were Ca19.9 (291 U/mL, n.v <40 U/mL), Ca125 (130.9 U/mL, n.v < 35 U/mL) and NSE (34 mcg/L, n.v < 12 mcg/L).

Since the patient had a history of PBC and a consequent higher risk of cholangiocarcinoma, magnetic resonance cholangiopancreatography (MRCP) and MRI with contrast were performed, excluding the presence of tumor lesions.

However, miRNA-based molecular prediction by PAMR reported with high probability (p=0.999) a biliary origin, which was consistent with clinical/pathological information available. Patient's treatment consisted in FOLFOX-4 regimen every 2 weeks up to 11 courses (partial response), followed by a maintenance treatment based on De Gramont regimen (FULCV) for 4 courses. Patient's conditions underwent a clinical worsening with the occurrence of pleural and abdominal

effusions. At November 2018 blood was collected for CTC analysis with CELLSEARCH and DEPArray system.

Treatment with Pazopanib, a novel multi-kinase inhibitor (VEGFR1, VEGFR2, VEGFR3, PDGFR, FGFR, c-Kit and c-Fms/CSF1R), was proposed due to the presence of *FGFR2* amplification (F1CDx testing). However, death occurred in June 2019 due to progressive disease and hepato-renal syndrome.

Case presentation: CB090

The second patient, CB090, was a 63-year-old woman admitted at Bologna University Hospital in February 2019 due to an oedema of the upper limb combined with weight loss, asthenia, nausea and cough.

Echo-Doppler scanning revealed a thrombosis in axillary, subclavian, basilica, internal and external jugular veins. Contrast-enhanced total body CT discovered the presence of supra- and infradiaphragmatic enlarged lymph nodes associated with oedema and diffuse suffusion of mesenteric and perivisceral tissues consistent with peritoneal carcinosis, confirmed also by a 18F-FDG-PET/CT scan.

CT-guided biopsy of a left paraortic lymph node led to a first pathological diagnosis of lymph node metastasis from poorly differentiated carcinoma with microglandular and signet-ring architecture. A first pathological hypothesis of a gastric origin was formulated based on IHC analysis performed on the lymph node metastasis reporting a $K7^+/CDX2^+/K20^-$/synaptophysin$^-$ profile.

Blood was collected for CTC analysis with CELLSEARCH/DEPArray and Parsortix systems at December 2019.

In May 2019, MSI and of HER2 status, the mutational status of *KRAS*, *NRAS*, *BRAF* and the methylation status of MGMT were all tested with negative results.

EGDS with ecoendoscopy identified a suspicious area in the ampulla of Vater. Histological examination confirmed the presence of poorly differentiated adenocarcinoma with mixed, micro-glandular pattern, and signet-ring cells and a lymphangitic infiltration of the mucosa of duodenal type. IHC staining confirmed the $K7^+/K20^-/CDX2^+$ profile and detected also a strong positivity to MUC1, MUC5AC, MUC6 and a complete negativity to MUC2, suggesting a likely bilio-pancreatic origin.

CB090 was treated with FOLFOX-4 regimen up to 9 cycles, obtaining an initial partial response. In October 2019 the patient experienced dyspnea and bilateral pleural effusion and underwent pleural drainage and cytological examination,

followed by videoassisted thoracoscopy and pleurodesis. IHC analysis of the cell block revealed a positivity for BEREP4, K7 and CDX2 and negativity for TTF1 and calretinin. Genetic analysis on pleural-derived cytological specimen, performed in the diagnostic setting, reported a pathogenic alteration in *FGFR2* gene (c.1150G>A, p. Gly384Arg, AF:33.4%) and the absence of gene amplification or fusion. Moreover, miRNA-based molecular prediction by both PAMR and LASSO classifiers pointed out with high probability (p>0.9) a gastrointestinal origin, which was concordant with clinical/pathological information available.

In December 2019, the patient started a weekly gemcitabine regimen as second-line treatment but after two weeks experienced uncontrolled ascitic effusion. Abdominal CT with and without contrast revealed progressive disease due to evident radiological signs of peritoneal carcinosis. The patient died in January 2020 caused by the further deterioration of patient's conditions.

Isolation and characterization of CTCs using CELLSEARCH and DEPArray system

We isolated CTCs from patients' peripheral blood by combining CELLSEARCH and DEPArray systems. For CB055, we captured with immunomagnetic beads up to 66 $CK^{+}/CD45^{-}/DAPI^{+}$ cells (not positive for PD-L1). The isolated cells were then loaded on DEPArray that retrieved n=12 CTCs. Three CTCs and 1 leukocyte as control, were subjected to single-cell whole genome amplification and sequencing.

On the other hand, from the peripheral blood of CB090, we captured 133 cells (defined as above) by CELLSEARCH and recovered with DEPArray system up to 34 putative CTCs (**Figure 12**). Single-cell whole genome amplification was performed on a total number of 32 CTCs and 4 leukocytes, and 10 CTCs passed quality control criteria for low-pass whole-genome analysis.

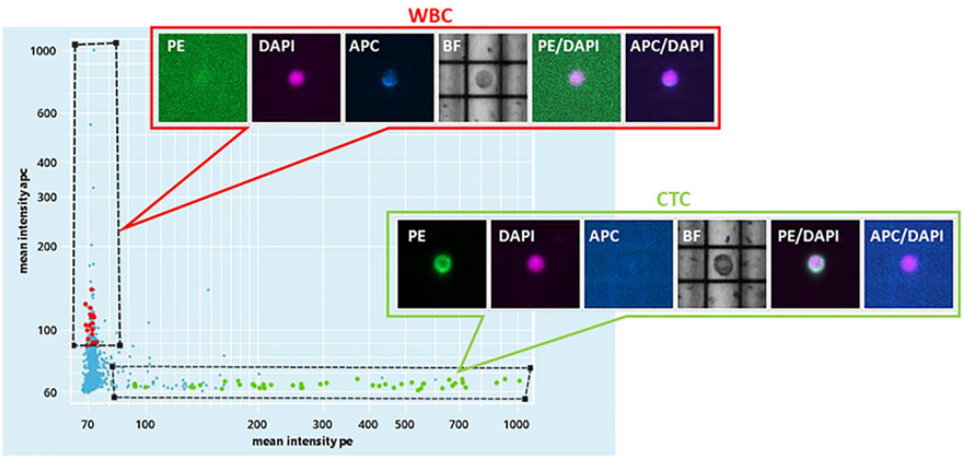

Figure 12. DEPArray NxT analysis. Scatter plot of the enriched blood sample of CB090. On the *x*-axis is reported the mean intensity fluorescence of PE channel, corresponding to the intensity of CK IF staining, while on the *y*-axis is represented the mean intensity of CD45-APC.

Isolation and characterization of CTCs using Parsortix system

In parallel, CB090 CTCs were also isolated by Parsortix system (ANGLE plc) a a 6.5 µm cassette. This system allowed a size-based capture of some EpCAM-positive and CD45-negative CTCs, along with EpCAM-positive and CD45-positive cells of unknown significance (**Figure 13**).

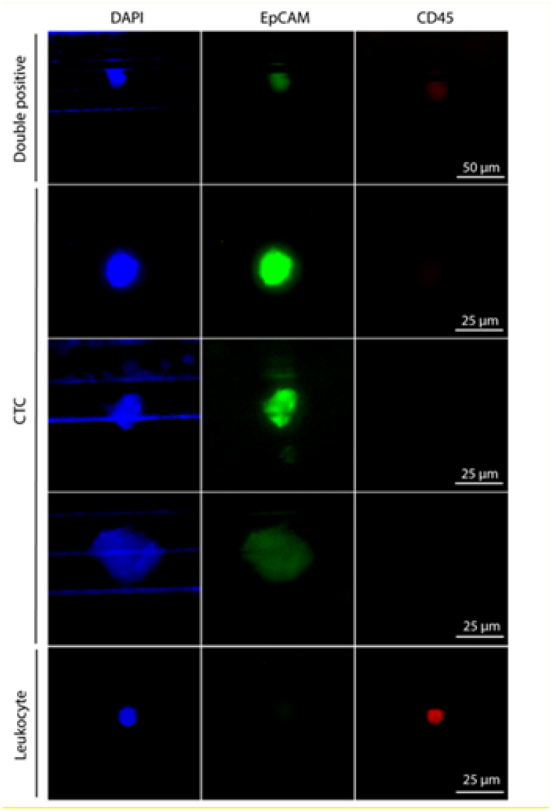

Figure 13. Staining of cells isolated with Parsortix system from CB090 patient.
Cells were forced to pass through 6.5 µm cassette in Parsortix system and then stained with different antibodies. Nuclei stained with DAPI (blue), CD45 (red) and EpCAM (green). From top to bottom: double-positive cells (CD45+, EpCAM+) of unknown significance, three exemplary CTCs (CD45-, EpCAM+), one exemplary leucocyte (CD45+, EpCAM-). Scale bar is 50 µm (double positive cells), 25 µm (CTCs and leucocyte).

Mutational analysis of CTCs, circulating tumor DNA and tumor

CB055

For CB055 F1CDx testing of the tumor tissue reported the amplification of *FGFR2* (Copy Number, CN: 49) and *CCNE1* (CN: 8) genes and a somatic alteration in *ARID1A* gene (p.R1276*) with an allele frequency (AF) of 7.3%. Moreover, F1CDx tumor mutational burden analysis reported 13 mutations/Mb, which corresponds to an intermediate burden.

Single-cell genome-wide characterization of CTCs (*Ampli*1 LowPass, Menarini Silicon Biosystems) revealed regions of sub-chromosomal loss, including loss of

heterozygosity (LOH) in a region that comprised the *APC* gene, and pattern of extensive gains, including *FGFR2* and *CCNE1* genes **(Figure 14)**.

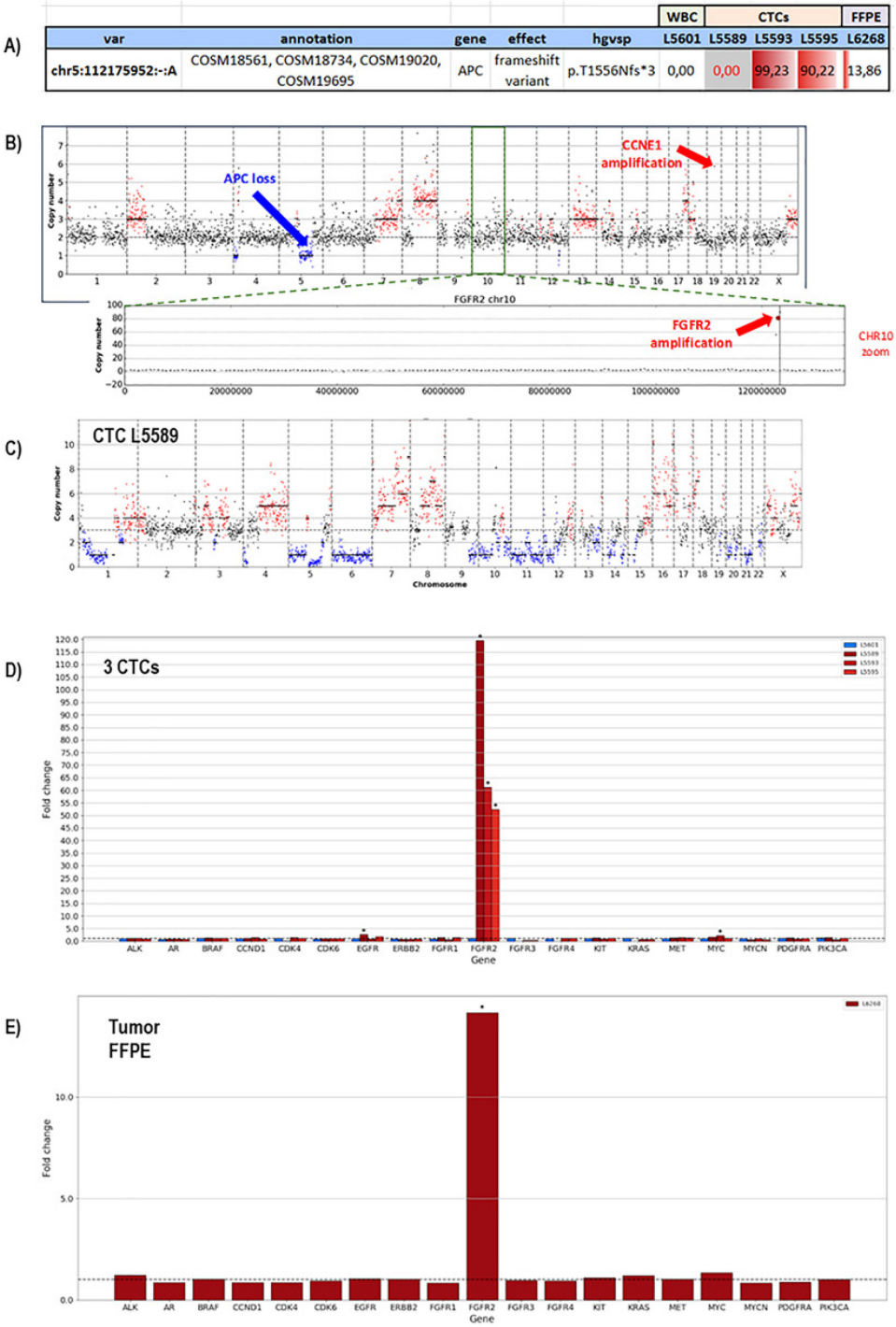

Figure 14. Whole-genome copy-number profiles and mutation analysis of CTCs and FFPE gDNA from CB055 patient. A) Identification of a homozygous somatic variant (absent in a single WBC), in 2 out of 3 CTCs in *APC* gene (p.T1556Nfs*3) also detected in the FFPE tissue; **B)** Patterns of copy-number profile of one exemplary CTC, in which is displayed a gene amplification in *FGFR2* and *CCNE1 (*red arrows), and LOH of a region of chromosome 5 that comprises the *APC* gene (blue arrow); **C)** LowPass profile of L5589 CTC showing the complete deletion of *APC* genomic region, in agreement with the lack of coverage in targeted sequencing (grayed out in A);

D) Copy number aberration (CNA) distribution showing *FGFR2* amplification in all 3 CTCs **D)** and tumor biopsy **E).**

These alterations were also observed on bulk tumor DNA. In addition, targeted sequencing (*Ampli*1 OncoSeek panel, Menarini Silicon Biosystems) found the *FGFR2* amplification in 3/3 CTCs and an inactivating homozygous somatic variant in *APC* gene (p.T1556Nfs*3, AF: 90.2-99.2%), previously not detected by F1CDx. Of note, this variant was identified (**Figure 14**) in 2 out of 3 analyzed CTCs, due to the lack of coverage of the third CTC in the APC region (L5589, greyed out), in agreement with the deletion detected by LowPass copy-number profile in this cell. As expected, the single-WBC control from the same sample (L5601) has a wild-type APC gene. Moreover, target sequencing (DEPArray OncoSeek panel, Menarini Silicon Biosystems) of tumor gDNA confirmed the *FGFR2* amplification and reported the same frameshift variant in *APC* (p.T1556Nfs*3), with a lower allele frequency (AF: 14%), probably ascribable to normal cells contamination. The single-copy homozygous variant observed in the *APC* gene in CTCs is compatible with a ~15% variant frequency in the presence of 70% of normal cells in FFPE. Of note, when setting the contamination parameter to 70% in the FFPE low-pass analysis, the bioinformatic pipeline was able to call the same gain/loss patterns observed in CTCs, which were otherwise undetectable without the correction factor (**Figure 15**).

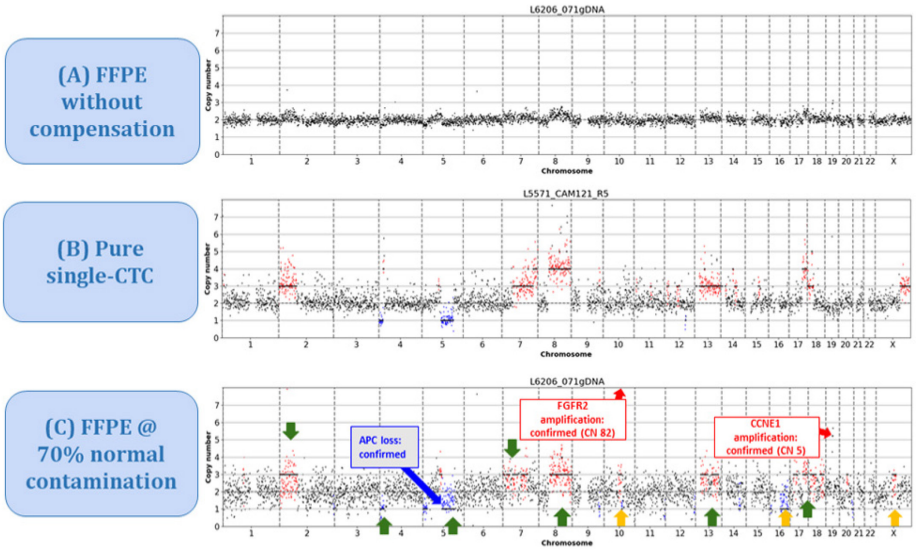

Figure 15. CB055: Low-Pass Whole Genome CNA profiles. A) FFPE without compensation: the signal produced by normal cells impact the copy-number alterations (CNA) call; **B)** Pure single-CTC: clear CNA calls with digital copy-number values (e.g., Chr8q n = 4). **C)** FFPE analysis set at 70% normal DNA contamination: signal is amplified and, among increased noise, some true positive CNA detected also in CTCs are called (green arrows) along with putative false-positives (orange arrows).

Target sequencing on CB055 using the SureSelect 92-gene custom panel was able to detect the APC insertion in bulk tumor DNA (AF:9.8%) and also in circulating cfDNAs analyzed at two different time points (ccfDNA at diagnosis, AF: 22.6% and ccfDNA at the progression of the disease, AF:29.9%). By using the tool *panelcn.MOPS* [223] and target sequencing data, we were able to confirm both *FGFR2* and *CCNE1* gene amplifications. The analysis reported a copy number alteration of 4 in both genes, which was the maximum value for this tool, but was not able to detect the *APC* loss. In addition, using this 92-gene panel we were able to confirm the alterations in *ARID1A* gene (p.R1276*), *ERBB3* (p.S128R), *KEAP1* (p.R135H), *PALB2* (p.P807S) and *NTRK1* (p.Q736X), previously identified by F1CDx, both in the tumor and in the ccfDNAs. All detected variants and CNVs observed in CB055 are listed in **Table 13** and **Table 14**.

In **Table 13** we reported other alterations detected by F1CDx in genes not included in the 92-gene custom panel and some variants detected in ccfDNAs not observed in the tumor. Specifically, we observed mutations in *ALK* (p.K434R, AF:7%), *EPHA5* (p.T473P, AF: 8.3%), *PTPRD* (p.Q1039H, AF:6.8%) and other mutations detected at lower frequency.

Table 13. Mutation report of tumor, ccfDNA and CTCs from CB055 [221].

										Pathogenicity Prediction								
Gene	**position**	**coding change**	**aminoacidic change**	**tumor FFPE** SureSelect 92-gene panel	**ccfDNA at diagnosis** SureSelect 92-gene panel	**ccfDNA at PD** SureSelect 92-gene panel	**tumor FFPE** Foundati-One CDx™	**CTCs** Ampli1™ OncoSeek and LowPass Kit	**tumor FFPE** DEPArray™ OncoSeek and LowPass Kit	**SIFT [225]**	**Polyphen2 [226]**	**LRT [227]**	**Mutation Taster [228]**	**Mutation Assessor [229]**	**FATHMM [230]**	**CADD [231]**	**VEST3 [232]**	**Actional genes / approved drugs (OncoKB [157])**
APC	chr5:112840254	4607insA	p.T1556fs*3	9.8%	22.6%	29.9%	ND	90.2%-99.2%	14.0%	-	-	-	-	-	-	-	-	-
ARID1A	chr1:26773456	c.C3826T	p.R1276X	8.9%	28.5%	22.8%	7.3%	NI	NI	-	-	-	DC	-	-	HF	-	-
ERBB3	chr12:56085144	c.C384A	p.S128R	9.8%	20.7%	20.4%	8.5%	NI	NI	-	PS	DT	DC	NFL	TL	BG	HF	-
NTRK1	chr1:156881475	C2206T	p.Q736X	8.3%	26.0%	22.3%	7.6%	NI	NI	DT	-	DT	DC	-	-	HF	-	-
KEAP1	chr19:10499630	c.G404A	p.R135H	9.0%	22.6%	19.5%	7.0%	NI	NI	TL	PD	DT	DC	NFL	TL	HF	HF	-
PALB2	chr16:23629735	C2419T	p.P807S	9.0%	19.1%	16.6%	6.0%	NI	NI	TL	BG	NL	PM	NFN	TL	BG	BG	Level 1/ Olaparib*
KMT2C	chr7:152265209	c.C1013T	p.S338L	ND	5.5%	4.3%	NI	NI	NI	TL	PD	U	DC	NFL	DG	BG	BG	-
KMT2C	chr7:152247975	c.C2459T	p.T820I	ND	1.1%	1.1%	NI	NI	NI	DT	PS	DT	DC	NFL	DG	HF	HF	-
PTPRD	chr9:8404618	c.T2908G	p.L970V	ND	5.1%	3.5%	NI	NI	NI	-	PD	DT	DC	FCM	TL	BG	HF	-
TP53	chr9:8404618	c.T316C	p.C106R	ND	2.1%	ND	ND	ND	ND	DT	PD	DT	DC	FCM	DG	HF	HF	Level 1-3 /no drug
TP53	chr17:7674251	c.G315C	p.M105I	ND	1.8%	ND	ND	ND	ND	DT	PD	DT	DC	FCM	DG	HF	HF	Level 1-3 /no drug
TP53	chr17:7674252	c.A355G	p.I119V	ND	1.7%	ND	ND	ND	ND	DT	PD	DT	DC	FCM	DG	HF	HF	Level 1-3 /no drug
TP53	chr17:7674212	c.C9G	p.C3W	ND	1.2%	ND	ND	ND	ND	DT	PD	DT	DC	FCM	DG	BG	HF	Level 1-3 /no drug
ALK	chr2:29328463	c.A1301G	p.K434R	ND	ND	7.0%	ND	OTR	OTR	-	BG	U	DC	NFN	TL	HF	BG	Level 1 / Lorlatinib, Brigatinib*
EPHA5	chr4:65420551	c.A1417C	p.T473P	ND	ND	8.3%	NI	NI	NI	TL	PS	NL	DC	NFL	TL	BG	HF	-

FAT1	chr4:186 707744	c.A2084C	p.N695T	ND	ND	2.6%	NI	NI	NI	TL	BG	NL	PM	NFN	TL	BG	BG	-
MGA	chr15:41 736268	c.T4004C	p.L1335 P	ND	ND	1.6%	NI	NI	NI	DT	PD	NL	DC	NFN	TL	BG	HF	-
KMT2C	chr7:152 248206	c.C2228T	p.P743L	ND	ND	1.6%	NI	NI	NI	TL	BG	NL	PM	NFN	DG	BG	BG	-
KMT2C	chr7:152 235897	c.C2689T	p.R897X	ND	ND	1.4%	NI	NI	NI	-	-	-	DC	-	-	HF	-	-
PTPRD	chr9:838 9280	c.A3117C	p.Q1039 H	ND	ND	6.8%	ND	NI	NI	-	PD	DT	DC	FCM	DG	BG	HF	-
ABL1	chr9:133 760790	c.C3113T	p. A1038V	NI	NI	NI	52.8%	OTR	OTR	-	BG	-	-	NFN	-	HF	-	-
GRM3	chr7:864 15916	c.C808T	p.R270C	NI	NI	NI	7.0%	NI	NI	-	PS	-	-	FCM	-	HF	-	-
MAP3K 1	chr5:561 77843	c.C2816G	p.S939C	NI	NI	NI	49.6%	NI	NI	-	BG	-	-	-	-	HF	-	-
MSH3	chr5:801 49992	c.A2857T	p.M953L	NI	NI	NI	46.1%	NI	NI	-	PD	-	-	FCM	-	HF	-	-
MSH6	chr2:480 28273	c.G3151A	p.V1051I	NI	NI	NI	50.0%	OTR	OTR	-	BG	-	-	NFL	-	BG	-	-
PTPRO	chr12:15 475691	c.G31T	p.A11S	NI	NI	NI	52.2%	NI	NI	-	BG	-	-	NFN	-	BG	-	-

*NI=not included; ND= not detected; OTR= out of target region; TL= tolerated; DT= deleterious; BG=benign; PD= Probably Damaging; PS= Possibly damaging; U= Unknown, NL=normal; DC= Disease causing; PM= polymorphism; NFN=non-functional neutral; NFL= non-functional low; FCM= functional medium; DG=Damaging; HF= harmful; * approved for cancer types different from CUP.*

Table 14. Copy number (CN) alterations detected in CB055 [221].

		SureSelect 92-gene panel	**SureSelect 92-gene panel**	**FoundatiOne CDx™**	**Ampli1™ OncoSeek and LowPass Kit**	**DEPArray™ OncoSeek and LowPass Kit**
Gene	***alteration***	***tumor FFPE (CN)***	***ccfDNAs (CN)***	***tumor FFPE (CN)***	***CTC (CN)***	***tumor FFPE (CN)***
FGFR2	amplification	>4	>4	49	>80	82
CCNE1	amplification	>4	>4	8	6	5
APC	loss/deletion	not detected	not detected	not detected	detected	detected

We then used droplet digital PCR technology to validate the *ARID1A* mutation (**Figure 16A**) and the *FGFR2* amplification (**Figure 16B**) in bulk tumor DNA and in longitudinally collected ccfDNAs at three different time points of disease evolution. Of note, *ARID1A* mutation frequency decreases at the second time point (Nov 2018), during FOLFOX-4 treatment, then it increases during the progression of the disease (May 2019). Moreover, the Variant Allele Frequency reported by target sequencing at the two analyzed time points (the first and the last) was found to be consistent with the Fractional Abundance (FA%= mutated copies/ mutated copies + wild type copies) measured by ddPCR. On the other hand, evaluating by ddPCR the *FGFR2* amplification in longitudinal ccfDNAs we observed a decreasing trend in term of copy number value during disease progression.

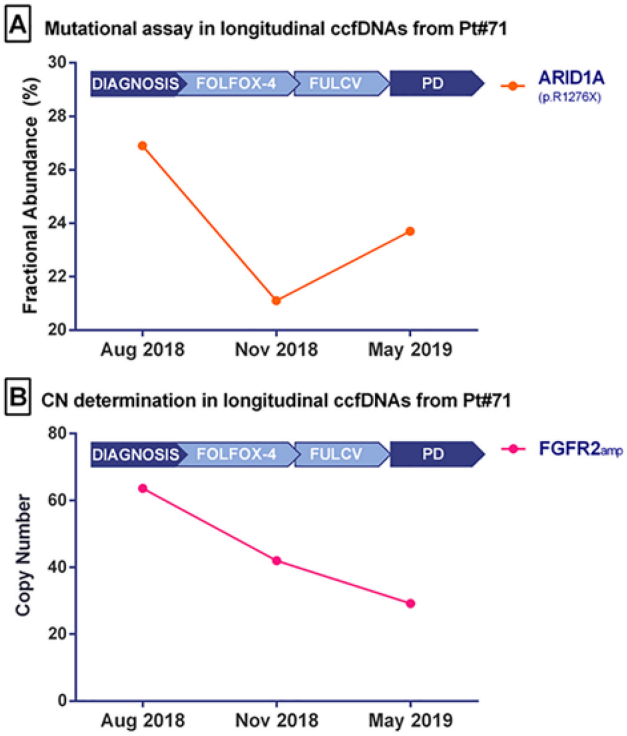

Figure 16. Kinetics of mutated fractional abundance and CNV in longitudinal ccfDNAs measured by Droplet Digital PCR in CB055. The graph in **A)** illustrates *ARID1A* mutated fractional abundance (p.R1276X) during patient's follow-up. This value decreases during treatment and later rises at the worsening of patient's conditions; on the contrary, when monitoring *FGFR2* amplification at the same time points, as reported in **B)**, we observed a decreasing trend from the diagnosis to disease progression.

CB090

For the second patient, single-cell whole genome analysis detected heterogeneous patterns of gain/loss in the 10 analyzed CTCs. However, amplification signals in the

regions of *ASPM* and *SEPT9* genes and deletion in the *FANCC* gene region was found to recur in all analyzed CTCs, as shown in the CNV profile clustering in **Figure 16**.

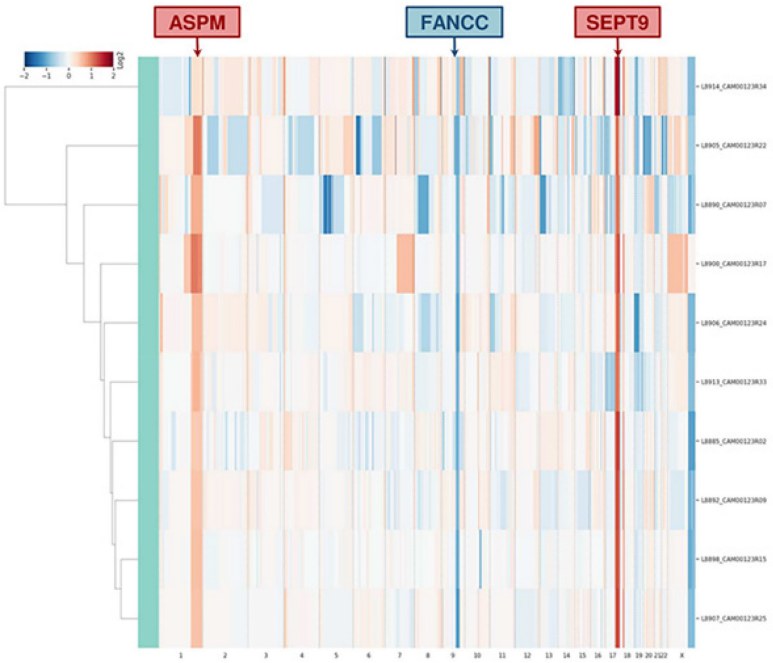

Figure 16. Clustering analysis of 10 CTCs based on single-cell whole genome copy number profile. The heatmap represents the relative copy number profiles of 10 different CTCs isolated through CELLSEARCH® and DEPArray™ technologies. Segments of gains or deletions are color-coded according their relative log2 copy number ratios (red=gain; blue=loss).

In this case, by targeted sequencing no somatic mutation in the regions covered by the panel (*Ampli*1 OncoSeek Panel, Menarini Silicon Biosystems) was found in both CTCs and bulk tumor DNA. However, target sequencing using the 92-gene custom panel, was able to identify 16 non-synonymous, somatic genetic alterations not detected in the normal cells in the ccfDNA and 8 in the tumor DNA (**Table 15**), including a deletion (I1485Rfs*19, AF: 32.6%) and a mutation (p. A1487V, AF: 33.2%) in *ARID1A* gene and a mutation in *FGFR2* gene (p.G272R, AF:19%). Of note, mutational analysis performed in the diagnostic setting (obtained with Oncomine Comprehensive Assay by Thermo Fisher Scientific) reported the same mutation in *FGFR2* gene; no amplifications or deletions were detected with SureSelect custom panel, but regions in proximity of *ASPM*, *SEPT9* and *FANCC* genes were not captured.

Table 15. Mutation report of tumor, ccfDNA and CTCs from CB090 [221].

							Pathogenicity Prediction								
Gene	**position**	**coding change**	**aminoacidic change**	**ccfDNA at diagnosis** SureSelect 92-gene panel	**tumor FFPE** SureSelect 92-gene panel	**CTCs** Ampli1™ OncoSeek and LowPass Kit	**SIFT [225]**	**Polyphen2 [226]**	**LRT [227]**	**Mutation Taster [228]**	**Assessor [229] Mutation**	**FATHMM [230]**	**CADD [231]**	**VEST3 [232]**	**Actional genes / drugs (OncoKB [157])**
ARID1A	chr1:26774679	c.4454_4457delTACA	p.I1485Rfs*19	32.6%	25%	NI	-	-	-	-	-	-	-	-	
ARID1A	chr1:26774687	c.C4460T	p.A1487V	33.2%	27%	NI	TL	BG	DT	DC	NFN	TL	BG	HF	
SPEN	chr1:15937527	c.G10391C	p.G3464A	31.0%	24%	NI	TL	BG	-	DC	NFL	TL	BG	BG	
PIK3CG	chr7:106872869	c.G2218A	p.E740K	22.0%	8%	NI	TL	BG	NL	DC	NFL	TL	BG	BG	
FGFR2	chr10:121515254	c.G814A	p.G272R	19.0%	NC	OTR	DT	PD	DT	DC	FCM	DG	HF	HF	Level 4 / Debio1347, AZD4547, BGJ398, Erdafitinib
KMT2D	chr12:49031574	c.G13131A	p.W4377X	17.0%	FLQ	NI	DT	-	NL	DC†	-	-	HF	-	Level 3 /no drug
FAT1	chr4:186621042	c.A5544C	p.Q1848H	13.1%	NC	NI	TL	PD	DT	DC	NFL	TL	BG	HF	
KMT2C	chr7:152265180	c.G1042A	p.D348N	9.8%	FLQ	NI	TL	PD	U	DC	NFL	DG	BG	BG	
KMT2C	chr7:152248140	c.A2294G	p.E765G	7.8%	ND	NI	TL	BG	DT	DC	NFL	DG	BG	BG	
KMT2C	chr7:152235876	c.C2710T	p.R904X	7.5%	FLQ	NI	TL	-	NL	DC	-	-	HF	-	
KMT2C	chr7:152235929	c.G2657A	p.R886H	6.8%	ND	NI	DT	PS	DT	DC	FCM	DG	HF	BG	
KMT2C	chr7:152235861	c.A2725G	p.R909G	6.6%	ND	NI	TL	PD	DT	DC	FCM	DG	BG	HF	
KMT2C	chr7:152248143	c.C2291T	p.S764F	4.9%	FLQ	NI	DT	BG	NL	DC	NFL	DG	BG	BG	
KMT2C	chr7:152265209	c.C1013T	p.S338L	4.9%	ND	NI	TL	PD	U	DC	NFL	DG	BG	BG	
KMT2C	chr7:152273774	c.G943A	p.G315S	4.9%	NC	NI	DT	PD	DT	DC	FCM	TL	HF	HF	
KMT2C	chr7:152247966	c.T2468C	p.I823T	2.3%	NC	NI	TL	PD	DT	DC	NFL	DG	BG	HF	
MGA	chr15:41767261	c.C8552T	p.P2851L	NC	5.20%	NI	TL	PS	NL	DC	NFN	DG	BG	BG	
KMT2C	chr7:152163206	c.C10371A	p.D3457E	NC	6.30%	NI	TL	BG	NL	DC	NFL	DG	BG	BG	
PTPRD	chr9:8521304	c.G904A	p.E302K	NC	5.00%	NI	-	PD	DT	DC	FCM	TL	HF	HF	
ASXL2	chr2:25743628	c.G1929T	p.Q643H	NC	5.10%	NI	DT	PD	NL	DC	FCM	TL	BG	BG	

NI=not included; NC= not covered region; FLQ= filtered for low quality; ND= not detected; OTR= out of target region; TL= tolerated; DT= deleterious; BG=benign; PD= Probably Damaging; PS= Possibly damaging; U= Unknown, NL=normal; DC= Disease causing; NFN=non-functional neutral; NFL= non-functional low; FCM= functional medium; DG=Damaging; HF= harmful; †known to be disease causing.

Discussion

The identification of the primary site in metastatic cancers strongly depends on clinical information and histology as well as immunohistochemical evaluations. Despite the notable progress in the pathological assessment of metastatic tumors observed in the last decade, this diagnostic workup is sometimes unsuccessful and a fraction of primaries remains undetected.

Metastatic cancers of unknown primary site (CUPs) are, by definition, advanced cancers whose site of origin is not detectable nor presumable, despite an intensive clinical and pathological diagnostic workup [5]. CUPs represent an enigma both at the biological and pathological level, and an important under-researched clinical problem.

Since treatment guidelines are tumor-oriented, without an indication of a primary site, CUP patients have limited treatment options and are mainly treated with poorly-effective platinum-based chemotherapy combinations, that consequently impact their chance of survival. An indication of a primary or the detection of driver/ actionable alterations, could potentially facilitate therapeutic decisions and give these patients access to more therapeutic options (i.e. target therapy or immunotherapy) and, potentially, prolong their life expectancy.

In the past decade, several molecular tests based on gene expression (GEP), microRNA or DNA methylation profile, were developed to predict CUPs' tissue-of-origin. The underlying premise for these molecular profiling assays is that metastatic tumors preserve core specific molecular signatures that match their primary site and can be used to infer their site-of-origin.

These assays, reporting prediction accuracies that range between 80-95%, are promising tools to obtain an indication of a primary site in CUP patients. Most recent NCCN occult primary guidelines support the employment of molecular profiling in the diagnostic workup of CUPs. However, at the present time, the clinical utility of these molecular assays needs to be further validated in large randomized phase III trials.

In this study, we designed a molecular assay able to evaluate in the same experiment the expression of 89 miRNAs in tumor FFPE samples using Droplet Digital PCR (ddPCR) technology, and allow CUP primary site prediction [215]. The 89-miRNA signature was determined combining two cancer-specific miRNA signatures previously published in two microarray-based studies [127,134]. Compared to large-scale technologies such as microarrays or sequencing, our

approach has several advantages, providing an on-demand quantification of a focused panel of miRNAs per sample, at an affordable cost and in a 2-days timeframe. Moreover, EvaGreen-based ddPCR technology allows miRNA absolute quantification without the requirement of standard curves, efficiency correction approaches or technical replicates [237], detecting target miRNAs at levels down to 1 copy/µL [238].

Our results confirmed the advantage of a miRNA-based instead of a GEP-based molecular assay, since we were able to successfully analyze all 159 FFPE samples of our cohort (100% success rate), with no excluded sample due to technical issues. We assessed the profile of 89 miRNAs in 159 FFPE samples, including 53 CUPs, obtaining a primary site prediction for all patients. Taking advantage of two different classifiers, our assay showed a good prediction accuracy when applied on metastatic cancers of known origin and showed highly consistent results on multiple metastases deriving from the same CUP patient, which provided an intrinsic validation of our prediction. Moreover, the employment of two predictive models allowed us to obtain stronger results when both systems pointed out to the same primary site. In addition, when available, clinical and/or histopathological hypothesis were found to be mostly consistent with molecular prediction.

Noteworthy, some CUPs molecularly predicted as LUAD, were negative for TTF1 by IHC staining, which characterizes a subgroup of LUAD with a poor prognosis [239].

Our results provide further evidence of the translational potential of a miRNA-based molecular testing in CUP patients. This assay was conceived to support the pathologist in narrowing the spectrum of possible primary sites, with no intention to replace IHC testing. CUPs tumor material is usually scarce due to the extensive use of tissue slices in the diagnostic setting for the immune-phenotypic characterization, so this approach could spare some material that could be used for genetic testing.

Moreover, when no pathological hypothesis can be formulated, the miRNA-based molecular prediction could assist the oncologists in their therapeutic choice, even though its clinical benefit in the diagnostic setting is still to be demonstrated.

Our ddPCR-based miRNA assay showed an accuracy comparable with other commercialized molecular tools, but overcomes some of their limits. In particular, our molecular classifiers cover a wide variety of primary cancers, among the most likely to be CUP's primary sites, and can discriminate up to 17 primary tumor subtypes. Other commercialized molecular assays are not able to cover such

number of tumor classes; i.e. the 10-genes qPCR assay (Veridex) can classify only six different tumor types. Our prediction results on CUPs are consistent with the frequency rates reported in post-mortem autopsy studies: lung (27%), pancreas (24%), liver or bile duct system (8%), kidney or adrenal (8%) or colon (7%) [233].
Three molecular assays were recently approved for CUP diagnostics in US: Pathwork Tissue of Origin Test (Pathwork Diagnostics), Cancer-TYPE ID (bioTheranostics), and miRview mets^2 (Rosetta Genomics). The first is a microarray-based system that is able to assess the GEP of 2000 genes and can discriminate up to 15 tumor types. CancerTYPE ID is GEP-based assay that assess the expression of a 92-gene signature by RT-qPCR and up to 30 tumor types. Finally, miRview mets^2 system, assessing the expression of 64 miRNAs by RT-qPCR, is able to recognize up to 26 tumor types.
However, contrarily to our molecular tool, these assays included primary tumors that have little or no connection with CUPs. Moreover, our assay has a 100% success rate and requires a 2-days working time, which is compatible with the application in the diagnostic setting and is considerably shorter compared to other commercial assays that present a turnaround time of 5-11 days.
In addition to being faster, targeted and cost-effective, our assay could be also combined with the analysis of druggable alterations, to assist the oncologists in the selection of the most suitable therapy.

To explore non-invasive approaches to genetically characterize CUP tumors, in the second part of this study we employed liquid biopsy approaches on two prospective CUP patients, using both circulating tumor cells (CTC) and circulating tumor DNA (ctDNA), with the aim to identify actionable genetic alterations.
The isolation and counting of CTCs are an important tool to evaluate aggressiveness and prognosis of different solid tumors. To date, CELLSEARCH system (Menarini Silicon Biosystems) is the only platform cleared by the FDA for the detection and count of CTCs in metastatic breast, prostate and colorectal cancer patients, while other technologies are at preclinical and clinical stages of development. CTC count, measured by CELLSEARCH System, was found to be a prognostic biomarker in different metastatic solid tumors, including breast [240], colorectal [241] and gastric cancers [242]. In addition, the detection of N≥1 CTCs using this system was found to predict early relapse and a worse outcome in chemo naïve patients with non-metastatic breast cancer [243].

Of note, in the blood of the two prospective CUP patients (CB055 and CB090), we detected a considerable number of CTCs that were successfully enriched and enumerated by CELLSEARCH and isolated with DEPArray for downstream molecular analysis (*Ampli*1 kit).

Parallelly to CELLSEARCH/DEPArray system, we also employed a label-free system, called Parsortix (ANGLE plc), that allows the enrichment of CTCs from peripheral blood of CUP patients by size-based selection. The advantage of this system is that there is no need of specific markers, which are difficult to select when dealing with CUP patients. Notably, some CTCs enriched with Parsortix system proved to be positive for EpCAM marker, therefore confirming their epithelial origin, although no genetic confirmation could be produced. We identified also CTCs that co-express hemato-epithelial markers ($EpCAM^+/CD45^+$). Such phenotype was previously observed in a subpopulation of epithelial ovarian cancer cells characterized by invasive mesenchymal features and the overexpression of genes related to drug resistance, anti-apoptosis and stemness [244]. The authors hypothesized that the origin of these $EpCAM^+/CD45^+$ CTCs was caused by a cell-cell fusion. The detection of these cells could be associated with CUP aggressive and highly-invasive phenotype, and deserves further investigations.

In this study, we used CTCs and ccfDNA for CUP genetic testing. Our results from two longitudinally monitored patients confirmed that CTCs and ctDNA can be isolated from the blood of CUP patients and are suitable for genetic characterization. Parallel analysis of FFPE tumor tissue was used to validate liquid biopsy results.

In addition, we employed and compared different targeted sequencing approaches applied to tumor FFPE tissue, CTCs and ccfDNA: specifically, FoundationOne CDx (Roche Foundation Medicine), OncoSeek Panels (*Ampli*1 and DEPArray OncoSeek Panels, Menarini Silicon Biosystems) and an in-house developed, Agilent SureSelect 92-gene panel, designed to include the most frequently mutated genes in CUPs and metastatic solid tumors. The comprehensive genomic analysis showed highly overlapping results among the different approaches, when the region was covered by all panels. Clearly, these three compared technologies include a different number of genes and hot spot regions, which explains the differential detection of some alterations.

Of note, both CUP patients presented actionable alterations in the *FGFR2* gene. Specifically, we detected a ~80-fold copy number amplification (CB055) and a gene mutation (CB090), both at tumor and CTC/ctDNA level, suggesting a potential

relevant role of this pathway in tumor progression and aggressiveness [245]. FGFR2 is a member of the fibroblast growth factor receptor family, whose signaling pathway is involved in cell proliferation, survival and migration [246]. *FGFR2* rearrangements have been observed in 10-15% of intrahepatic cholangiocarcinomas [247,248], amplifications in up to 15% of gastric cancers [249,250] and point mutations in about 10% of endometrial cancers [251]; in addition, *FGFR2* amplification has also a prognostic value in gastric cancers [250,252]. Noteworthy, in longitudinal ccfDNAs from CB055 we observed a progressive reduction of *FGFR2* CN during treatment, which reflects the initial partial response. Due to the chemo-naïve status of these patients and the unavailability of clinical trials testing multi-kinase or FGFR2 inhibitors in CUPs, they were treated according to the current guidelines with platinum salts and fluoropyrimidines.

CB055 had also an amplification in *CCNE1* (Cyclin E1) gene. Alterations in this gene were found to be associated with an unfavorable outcome in ovarian and triple-negative breast cancers, probably due to its involvement in the occurrence of chemoresistance [253,254].

Another common genetic characteristic of these two CUP cases, was the presence of point mutation in *ARID1A* gene, a tumor suppressor gene that encodes the DNA-binding subunit of the chromatin remodeler complex SWI/SNF [255]. Despite currently not druggable, this gene is the most frequently mutated chromatin regulator across all human cancers [256] and its loss compromises RNA polymerase II (RNAPII) pausing. The altered activity of RNAPII has oncogenic effects through the dysregulation of transcription of active genes, such as *TP53* and *ESR1* [255]. Since mutations in this gene were observed in both the reported cases, further investigation is required to evaluate a potential role of ARID1A in regulating the aggressiveness of cancers of unknown primary. Of note, when *ARID1A* mutation (p.R1276*) was assessed in longitudinal ccfDNAs of CB055, we observed that its Fractional Abundance decreased during treatment and then increases at disease progression, suggesting its potential involvement in disease worsening. Moreover, in the timepoints were the Fractional Abundance of $ARID1A_{mut}$ in ccfDNA was higher, there was also the emergence of novel mutations in *TP53, ALK, EPHA5, FAT1* and *MGA* genes, not detected in the tumor biopsy. Recent in vitro experiments reported evidence that *ARID1A* deficiency is associated with DNA replication stress [257], impaired checkpoint activation and repair of DNA double-strand breaks (DSBs) [258]. ARID1A in normal conditions is responsible for the chromatin

recruitment of MSH2 during DNA replication and the promotion of mismatch repair (MMR); however, its loss is associated with MMR disfunction and consequently, an increase in mutagenesis and mutational load [259]. Consequently, ARID1A loss has been investigated as a biomarker of response to anti-PD-1/PD-L1 immunotherapy treatment [260]; of note, regardless of microsatellite and tumor mutational burden status, alterations in ARID1A were found to be associated with a longer progression-free survival in patients treated with anti-PD-1/PD-L1 immunotherapy. Alterations in *ARID1A* gene was identified in 15% of CUP patients (MSK impact study [152]) for which immune checkpoint inhibitors could represent a valid therapeutic option. In these two CUP patients the accumulation of novel mutations observed in ccfDNAs could be explained by ARID1A-dependent accumulation of genetic alterations or, alternatively, by the mutagenic activity of oxaliplatin (L-OHP), a DNA-alkylating-like agent, which is included in FOLFOX-4 chemotherapy regimen [261,262]. This chemotherapy combination, that includes L-OHP along with Leucovorin and 5-fluorouracil (5-FU), is commonly used to treat metastatic colorectal cancer [263], but also CUP patients with $K20^{+}/K7^{-}/CDX2^{+}$ immunohistochemical signature.

Conclusions

Riding the wave of the rising consensus to recommend molecular profiling for tissue-of-origin prediction, either based on coding or non-coding transcripts or epigenetic modifications, in this study we demonstrated that miRNA expression profiling of CUPs is feasible in FFPE samples and has a remarkable translational potential.

Our assay can be performed on-demand, in association with standard diagnostic workup and genetic testing, to offer valuable indication about the possible primary site and thereby supporting treatment decisions.

A liquid biopsy approach, either using CTCs or ccfDNA, on CUP patients for genetic testing could help detecting druggable alterations taking advantage of its non-invasiveness and its potential application for longitudinal monitoring. This proof-of-concept study highlighted the importance of liquid biopsy genetic testing in CUPs for the identification of driver/actionable alterations and the relevance of using CUP-dedicated NGS panels. Molecular diagnostics and genetic profiling, are of the outmost importance to offer these patients novel therapeutic options and potentially improve their survival and should become the standard in CUP management.

References

1. Steeg, P.S. *Nat Med*, 2006. **12**(8): p. 895-904.
2. Chaffer, C.L. and R.A. Weinberg. *Science*, 2011. **331**(6024): p. 1559-64.
3. Hanahan, D. and R.A. Weinberg. *Cell*, 2011. **144**(5): p. 646-74.
4. Fares, J., et al. *Signal Transduct Target Ther*, 2020. **5**(1): p. 28.
5. Fizazi, K., et al. *Ann Oncol*, 2015. **26 Suppl 5**: p. v133-8.
6. Lampe, I. and G. Santin. *Univ Hosp Bull*, 1946. **12**(11): p. 106.
7. Laprovitera, N.R., M.; Ambrosini, E.; Klec, C.; Pichler, M.; Ferracin M. *Cancers*, 2021. **13**(3): p. 451.
8. Pavlidis, N. *Ann Oncol*, 2003. **14 Suppl 3**: p. iii11-8.
9. Abbruzzese, J.L., et al. *J Clin Oncol*, 1994. **12**(6): p. 1272-80.
10. Kirsten, F., et al. *Q J Med*, 1987. **62**(238): p. 143-61.
11. Le Chevalier, T., et al. *Arch Intern Med*, 1988. **148**(9): p. 2035-9.
12. Golfinopoulos, V., et al. *Cancer Treat Rev*, 2009. **35**(7): p. 570-3.
13. Tomuleasa, C., et al. *J Gastrointestin Liver Dis*, 2017. **26**(1): p. 69-79.
14. Kolling, S., et al. *Front Oncol*, 2019. **9**: p. 1546.
15. Rassy, E. and N. Pavlidis. *Cancer Epidemiol*, 2019. **61**: p. 139-141.
16. Dyrvig, A.K., et al. *Medicine (Baltimore)*, 2017. **96**(16): p. e6693.
17. Brustugun, O.T. and A. Helland. *Acta Oncol*, 2014. **53**(1): p. 134-7.
18. Binder, C., et al. *Cancer Medicine*, 2018. **7**(9): p. 4814-4824.
19. Schroten-Loef, C., et al. *European Journal of Cancer*, 2018. **101**: p. 77-86.
20. Shu, X., et al. *European Journal of Cancer Prevention*, 2012. **21**(3): p. 281-288.
21. Brewster, D.H., et al. *Cancer Epidemiology*, 2014. **38**(3): p. 227-234.
22. Luke, C., et al. *Aust N Z J Public Health*, 2008. **32**(4): p. 383-9.
23. Mnatsakanyan, E., et al. *Cancer Causes Control*, 2014. **25**(6): p. 747-57.
24. Urban, D., et al. *Br J Cancer*, 2013. **109**(5): p. 1318-24.
25. Pavlidis, N. and K. Fizazi. *Crit Rev Oncol Hematol*, 2009. **69**(3): p. 271-8.
26. Pavlidis, N., E. Rassy, and J. Smith-Gagen. *Int J Cancer*, 2020. **146**(6): p. 1490-1498.
27. Raghav, K., et al. *PLoS One*, 2016. **11**(5): p. e0154985.
28. Rassy, E., J. Kattan, and N. Pavlidis. *Int J Clin Oncol*, 2019. **24**(10): p. 1328-1331.
29. Kaaks, R., et al. *Int J Cancer*, 2014. **135**(10): p. 2475-81.
30. Rassy, E., P. Nicolai, and N. Pavlidis. *Head Neck*, 2019. **41**(10): p. 3700-3711.
31. Conway, A.M., et al. *Br J Cancer*, 2019. **120**(2): p. 141-153.
32. Davalos, V. and M. Esteller. *EMBO Mol Med*, 2020. **12**(7): p. e12685.
33. Scadden, D.T. *Nature*, 2006. **441**(7097): p. 1075-9.
34. Lopez-Lazaro, M. *Oncoscience*, 2015. **2**(5): p. 467-75.
35. Joosse, S.A. and K. Pantel. *Cancer Metastasis Rev*, 2016. **35**(1): p. 41-8.
36. Rassy, E., T. Assi, and N. Pavlidis. *Br J Cancer*, 2020. **122**(8): p. 1124-1132.
37. Vikesa, J., et al. *BMC Cancer*, 2015. **15**: p. 151.
38. Chitty, J.L., et al. *F1000Res*, 2018. **7**.
39. Kaplan, R.N., et al. *Nature*, 2005. **438**(7069): p. 820-7.
40. Tarin, D. *Cancer Microenviron*, 2012. **5**(2): p. 95-112.
41. Suzuki, M., et al. *Am J Pathol*, 2006. **169**(2): p. 673-81.
42. Speel, E.J., et al. *Int J Cancer*, 2008. **123**(6): p. 1292-300.
43. Benvenuti, S., et al. *EMBO Mol Med*, 2020. **12**(7): p. e11756.
44. Stoyianni, A., et al. *Clin Transl Oncol*, 2014. **16**(8): p. 725-31.
45. Stoyianni, A., et al. *Anticancer Res*, 2012. **32**(4): p. 1273-81.
46. Sampieri, K. and R. Fodde. *Semin Cancer Biol*, 2012. **22**(3): p. 187-93.
47. Losa, F., et al. *Clin Transl Oncol*, 2018. **20**(1): p. 89-96.
48. NCCN. *Occult Primary (Version 2.2020)*. Available from: https://www.nccn.org/professionals/physician_gls/pdf/occult_blocks.pdf.
49. Pentheroudakis, G., E. Briasoulis, and N. Pavlidis. *Oncologist*, 2007. **12**(4): p. 418-25.
50. Stewart, J.F., et al. *Br Med J*, 1979. **1**(6177): p. 1530-3.

51. Steckel, R.J. and A.R. Kagan. *Radiology*, 1980. **134**(2): p. 367-9.
52. Snyder, R.D., G.M. Mavligit, and M. Valdivieso. *Med Pediatr Oncol*, 1979. **6**(4): p. 389-94.
53. Nystrom, J.S., et al. *Semin Oncol*, 1977. **4**(1): p. 53-8.
54. Didolkar, M.S., et al. *Ann Surg*, 1977. **186**(5): p. 625-30.
55. Mayordomo, J.I., et al. *Tumori*, 1993. **79**(5): p. 321-4.
56. Maiche, A.G. *Am J Clin Oncol*, 1993. **16**(1): p. 26-9.
57. Hamilton, C.S. and A.O. Langlands. *Int J Radiat Oncol Biol Phys*, 1987. **13**(10): p. 1497-503.
58. Blaszyk, H., A. Hartmann, and J. Bjornsson. *APMIS*, 2003. **111**(12): p. 1089-94.
59. Al-Brahim, N., et al. *Ann Diagn Pathol*, 2005. **9**(2): p. 77-80.
60. Jordan, W.E., 3rd and R.A. Shildt. *Cancer*, 1985. **55**(4): p. 857-60.
61. Ramskold, D., et al. *PLoS Comput Biol*, 2009. **5**(12): p. e1000598.
62. Gevaert, O., et al. *BMC Med Genomics*, 2009. **2**: p. 69.
63. Wang, N.P.Z., S.; Zarbo, R.J.; Bacchi, C.E.; Gown, A.M. *Applied Immuno-histochemistry & Molecular Morphology*, 1995. **3**: p. 99-107.
64. Chu, P., E. Wu, and L.M. Weiss. *Mod Pathol*, 2000. **13**(9): p. 962-72.
65. Schweizer, J., et al. *J Cell Biol*, 2006. **174**(2): p. 169-74.
66. Duval, J.V., L. Savas, and B.F. Banner. *Arch Pathol Lab Med*, 2000. **124**(8): p. 1196-200.
67. Wildi, S., et al. *Clin Cancer Res*, 1999. **5**(10): p. 2840-7.
68. Shimonishi, T., K. Miyazaki, and Y. Nakanuma. *Histopathology*, 2000. **37**(1): p. 55-63.
69. Selves, J., et al. *Cancers (Basel)*, 2018. **10**(4).
70. Kato, S., et al. *Trends Cancer*, 2021. **7**(5): p. 465-477.
71. Kandalaft, P.L. and A.M. Gown. *Arch Pathol Lab Med*, 2016. **140**(6): p. 508-23.
72. Lin, F. and H. Liu. *Arch Pathol Lab Med*, 2014. **138**(12): p. 1583-610.
73. Massard, C., Y. Loriot, and K. Fizazi. *Nat Rev Clin Oncol*, 2011. **8**(12): p. 701-10.
74. Hainsworth, J.D. and K. Fizazi. *Semin Oncol*, 2009. **36**(1): p. 44-51.
75. Greco, F.A. and N. Pavlidis. *Semin Oncol*, 2009. **36**(1): p. 65-74.
76. Culine, S., et al. *J Clin Oncol*, 2002. **20**(24): p. 4679-83.
77. Zarkavelis, G., D. Mauri, and G. Pentheroudakis. *ESMO Open*, 2019. **4**(Suppl 2): p. e000502.
78. Warner, E., et al. *Br J Cancer*, 1998. **77**(12): p. 2376-80.
79. Pentheroudakis, G., et al. *Acta Oncol*, 2008. **47**(6): p. 1148-55.
80. Briasoulis, E., et al. *Cancer Chemother Pharmacol*, 2008. **62**(2): p. 277-84.
81. Huebner, G., et al. *Br J Cancer*, 2009. **100**(1): p. 44-9.
82. Hainsworth, J.D., et al. *Cancer*, 2015. **121**(10): p. 1654-61.
83. Gross-Goupil, M., et al. *Eur J Cancer*, 2012. **48**(5): p. 721-7.
84. Ramaswamy, S., et al. *Proc Natl Acad Sci U S A*, 2001. **98**(26): p. 15149-54.
85. Golub, T.R., et al. *Science*, 1999. **286**(5439): p. 531-7.
86. Su, A.I., et al. *Cancer Res*, 2001. **61**(20): p. 7388-93.
87. Monzon, F.A., et al. *J Clin Oncol*, 2009. **27**(15): p. 2503-8.
88. Pillai, R., et al. *J Mol Diagn*, 2011. **13**(1): p. 48-56.
89. Dumur, C.I., et al. *J Mol Diagn*, 2008. **10**(1): p. 67-77.
90. Handorf, C.R., et al. *Am J Surg Pathol*, 2013. **37**(7): p. 1067-75.
91. Monzon, F.A., et al. *Diagn Pathol*, 2010. **5**: p. 3.
92. Horlings, H.M., et al. *J Clin Oncol*, 2008. **26**(27): p. 4435-41.
93. Yoon, H.H., et al. *Ann Oncol*, 2016. **27**(2): p. 339-44.
94. Tothill, R.W., et al. *Pathology*, 2015. **47**(1): p. 7-12.
95. Ma, X.J., et al. *Arch Pathol Lab Med*, 2006. **130**(4): p. 465-73.
96. Kerr, S.E., et al. *Clin Cancer Res*, 2012. **18**(14): p. 3952-60.
97. Weiss, L.M., et al. *J Mol Diagn*, 2013. **15**(2): p. 263-9.
98. Hainsworth, J.D., et al. *J Clin Oncol*, 2013. **31**(2): p. 217-23.
99. Talantov, D., et al. *J Mol Diagn*, 2006. **8**(3): p. 320-9.
100. Greco, F.A., et al. *J Natl Cancer Inst*, 2013. **105**(11): p. 782-90.
101. Varadhachary, G.R., et al. *J Clin Oncol*, 2008. **26**(27): p. 4442-8.

102. Grewal, J.K., et al. *JAMA Netw Open*, 2019. **2**(4): p. e192597.
103. Greco, F.A., et al. *Mol Diagn Ther*, 2015. **19**(2): p. 91-7.
104. Hainsworth, J.D. and F. Anthony Greco. *Drugs Real World Outcomes*, 2016. **3**: p. 115-120.
105. AC, D.E.M., et al. *Oncol Lett*, 2015. **10**(4): p. 2657-2661.
106. Hayashi, H., et al. *J Clin Oncol*, 2019. **37**(7): p. 570-579.
107. Kurahashi, I., et al. *PLoS One*, 2013. **8**(5): p. e63249.
108. Fizazi, K.M., A.; Penel, N.; Baciarello, G.; Allouache, D.; Daugaard, G.; Van de Wouw, A.; Soler, G.; Vauleon, E.; Chaigneau, L.; Jansen, R.; Losa Gaspa, F.; Morales Barrera, R.; Balana, C.; Tosi, D.; Chauffert, B.; Schnabel, C.A.; Martineau, G.; Cline, S.; Borget, I. *Annals of Oncology*, 2019.
109. Ettinger, D.S., et al. *J Natl Compr Canc Netw*, 2014. **12**(7): p. 969-74.
110. Laprovitera, N., et al. *Cancers (Basel)*, 2021. **13**(3).
111. Tothill, R.W., et al. *Cancer Res*, 2005. **65**(10): p. 4031-40.
112. Bridgewater, J., et al. *Br J Cancer*, 2008. **98**(8): p. 1425-30.
113. Ferracin, M., et al. *J Pathol*, 2011. **225**(1): p. 43-53.
114. Mueller, W.C., et al. *Oncologist*, 2011. **16**(2): p. 165-74.
115. Varadhachary, G.R., et al. *Clin Cancer Res*, 2011. **17**(12): p. 4063-70.
116. Fernandez, A.F., et al. *Genome Res*, 2012. **22**(2): p. 407-19.
117. Pentheroudakis, G., et al. *Mol Cancer*, 2013. **12**: p. 57.
118. Moran, S., et al. *Lancet Oncol*, 2016. **17**(10): p. 1386-1395.
119. Opitz, L., et al. *BMC Med Genomics*, 2010. **3**: p. 36.
120. Cronin, M., et al. *Am J Pathol*, 2004. **164**(1): p. 35-42.
121. Masuda, N., et al. *Nucleic Acids Res*, 1999. **27**(22): p. 4436-43.
122. von Ahlfen, S., et al. *PLoS One*, 2007. **2**(12): p. e1261.
123. Liu, A., et al. *Int J Clin Exp Pathol*, 2009. **2**(6): p. 519-27.
124. Peiro-Chova, L., et al. *Virchows Arch*, 2013. **463**(6): p. 765-74.
125. Doleshal, M., et al. *J Mol Diagn*, 2008. **10**(3): p. 203-11.
126. Li, J., et al. *BMC Biotechnol*, 2007. **7**: p. 36.
127. Ferracin, M., et al. *Journal of Pathology*, 2011. **225**(1): p. 43-53.
128. Meiri, E., et al. *Oncologist*, 2012. **17**(6): p. 801-12.
129. Volinia, S., et al. *Proc Natl Acad Sci U S A*, 2006. **103**(7): p. 2257-61.
130. Lu, J., et al. *Nature*, 2005. **435**(7043): p. 834-8.
131. Tili, E., et al. *Carcinogenesis*, 2010. **31**(9): p. 1561-1566.
132. Farazi, T.A., et al. *J Pathol*, 2011. **223**(2): p. 102-15.
133. Sokilde, R., et al. *J Mol Diagn*, 2014. **16**(1): p. 106-15.
134. Rosenfeld, N., et al. *Nat Biotechnol*, 2008. **26**(4): p. 462-9.
135. Rosenwald, S., et al. *Mod Pathol*, 2010. **23**(6): p. 814-23.
136. Pentheroudakis, G., et al. *Clin Exp Metastasis*, 2013. **30**(4): p. 431-9.
137. Rigakos, G., et al. *Mol Clin Oncol*, 2016. **5**(2): p. 263-266.
138. Tang, W., et al. *Bioinformatics*, 2018. **34**(3): p. 398-406.
139. Ross, J.S., et al. *JAMA Oncol*, 2015. **1**(1): p. 40-49.
140. Clynick, B., et al. *J Transl Med*, 2018. **16**(1): p. 185.
141. Varghese, A.M., et al. *Ann Oncol*, 2017. **28**(12): p. 3015-3021.
142. Tothill, R.W., et al. *J Pathol*, 2013. **231**(4): p. 413-23.
143. Gatalica, Z., et al. *Oncotarget*, 2014. **5**(23): p. 12440-7.
144. Subbiah, I.M., et al. *Oncoscience*, 2017. **4**(5-6): p. 47-56.
145. Stella, G.M., et al. *Hum Mutat*, 2011. **32**(1): p. 44-50.
146. Pentheroudakis, G., et al. *Clin Exp Metastasis*, 2014. **31**(7): p. 761-9.
147. Cancer Genome Atlas Research, N., et al. *Nat Genet*, 2013. **45**(10): p. 1113-20.
148. Hoadley, K.A., et al. *Cell*, 2018. **173**(2): p. 291-304 e6.
149. Hoadley, K.A., et al. *Cell*, 2014. **158**(4): p. 929-944.
150. Tsimberidou, A.M., et al. *JCO Precis Oncol*, 2017. **2017**.
151. Ross, J.S., et al. *JAMA Oncol*, 2015. **1**(1): p. 40-9.

152. Zehir, A., et al. *Nat Med*, 2017. **23**(6): p. 703-713.
153. Cerami, E., et al. *Cancer Discov*, 2012. **2**(5): p. 401-4.
154. Piscuoglio, S., et al. *J Pathol*, 2016. **238**(4): p. 508-18.
155. Huang, F.W., et al. *Science*, 2013. **339**(6122): p. 957-9.
156. Consortium, A.P.G. *Cancer Discov*, 2017. **7**(8): p. 818-831.
157. Chakravarty, D., et al. *JCO Precision Oncology*, 2017(1): p. 1-16.
158. Prahallad, A., et al. *Nature*, 2012. **483**(7387): p. 100-3.
159. Le Tourneau, C., et al. *Lancet Oncol*, 2015. **16**(13): p. 1324-34.
160. Massard, C., et al. *Cancer Discov*, 2017. **7**(6): p. 586-595.
161. Tredan, O., et al. *Ann Oncol*, 2019. **30**(5): p. 757-765.
162. Groschel, S., et al. *Cold Spring Harb Mol Case Stud*, 2016. **2**(6): p. a001180.
163. Palma, N.A., et al. *Case Rep Oncol*, 2014. **7**(2): p. 503-8.
164. Tan, D.S., et al. *J Clin Oncol*, 2013. **31**(14): p. e237-9.
165. Chung, J.H., et al. *Case Rep Oncol*, 2014. **7**(3): p. 628-32.
166. Hainsworth, J.D., et al. *Clin Colorectal Cancer*, 2012. **11**(2): p. 112-8.
167. Mileshkin, L.R., et al. 2019. **37**(15_suppl): p. 3072-3072.
168. Haratani, K., et al. *J Immunother Cancer*, 2019. **7**(1): p. 251.
169. Gatalica, Z., et al. *Eur J Cancer*, 2018. **94**: p. 179-186.
170. Chalmers, Z.R., et al. *Genome Med*, 2017. **9**(1): p. 34.
171. Le, D.T., et al. *N Engl J Med*, 2015. **372**(26): p. 2509-20.
172. Le, D.T., et al. *Science*, 2017. **357**(6349): p. 409-413.
173. Garon, E.B., et al. *N Engl J Med*, 2015. **372**(21): p. 2018-28.
174. Eggermont, A.M.M., et al. *N Engl J Med*, 2018. **378**(19): p. 1789-1801.
175. Kato, S., et al. *Cancer Res*, 2017. **77**(16): p. 4238-4246.
176. Tanizaki, J.Y., K.; Akiyoshi, K.; Minami, H.; Ueda, H.; Takiguchi, H.; Nakayama Kondoh, C.; Segawa, Y.; Takahashi, S.; Iwamoto, Y.; Kidera, Y.; Fukuoka, K.; Nakamura, Y.; Chiba, Y.; Nishio, K.; Nakagawa, K.; Hayashi, H. *Journal of Clinical Oncology*, 2020. **38**(15_suppl): p. 106-106.
177. Ross, J.S.S., E.S.; Moch, H.; Mileshkin, L.; Baciarello, G.; Losa, F.; Beringer, A.; Thomas, M.; Foser, S.; Elvin, J.; Danziger, N.; Ngo, N.; Tse, J.Y.; Killian, K.; Jin, D.X.; Gay, L.M.; Kramer, A. *Annals of Oncology*, 2019.
178. Loffler, H., et al. *Oncotarget*, 2016. **7**(28): p. 44322-44329.
179. Marquard, A.M., et al. *BMC Med Genomics*, 2015. **8**: p. 58.
180. Miyamoto, D.T., et al. *Cancer Discov*, 2012. **2**(11): p. 995-1003.
181. Jiao, W., et al. *Nat Commun*, 2020. **11**(1): p. 728.
182. Bagge RO, D.A., Karlsson J, Alaei-Mahabadi B, Einarsdottir BO, Jespersen H, Lindberg MF,Muth A, Nilsson LM, Persson M, Svensson JB, Söderberg EBV, de Krijger RR, Nilsson O, Larsson E, Stenman G, Nilsson JA. *JCO Precision Oncology*, 2018.
183. Penson, A., et al. *JAMA Oncol*, 2019.
184. Gerlinger, M., et al. *N Engl J Med*, 2012. **366**(10): p. 883-892.
185. Wong, S.Q., et al. *BMC Med Genomics*, 2014. **7**: p. 23.
186. El Rassy, E., H. Khaled, and N. Pavlidis. *Eur J Cancer*, 2018. **105**: p. 28-32.
187. Klein, E.A., et al. 2018. **36**(15_suppl): p. 12021-12021.
188. Lu, S.H., et al. *Cancer Biol Ther*, 2016. **17**(4): p. 430-8.
189. Lehmann-Werman, R., et al. *Proc Natl Acad Sci U S A*, 2016. **113**(13): p. E1826-34.
190. Best, M.G., et al. *Cancer Cell*, 2015. **28**(5): p. 666-676.
191. Fece de la Cruz, F. and R.B. Corcoran. *Ann Oncol*, 2018. **29**(6): p. 1351-1353.
192. Schwaederle, M., et al. *Oncotarget*, 2016. **7**(9): p. 9707-17.
193. Hoshino, A., et al. *Cell*, 2020. **182**(4): p. 1044-1061 e18.
194. Kang, S., et al. *Genome Biol*, 2017. **18**(1): p. 53.
195. Matthew, E.M., et al. *Oncotarget*, 2016. **7**(4): p. 3662-76.
196. Ignatiadis, M. and S.J. Dawson. *Ann Oncol*, 2014. **25**(12): p. 2304-2313.
197. Pairawan, S., et al. *Clin Cancer Res*, 2020. **26**(8): p. 1924-1931.

198. Lebofsky, R., et al. *Mol Oncol*, 2015. **9**(4): p. 783-90.
199. Jovelet, C., et al. *Clin Cancer Res*, 2016. **22**(12): p. 2960-8.
200. Ikeda, S., et al. *J Hematol Oncol*, 2018. **11**(1): p. 76.
201. Aceto, N., et al. *Cell*, 2014. **158**(5): p. 1110-1122.
202. Joosse, S.A., T.M. Gorges, and K. Pantel. *EMBO Mol Med*, 2015. **7**(1): p. 1-11.
203. Riethdorf, S., et al. *Clin Cancer Res*, 2007. **13**(3): p. 920-8.
204. Sparano, J., et al. *JAMA Oncol*, 2018. **4**(12): p. 1700-1706.
205. Yu, M., et al. *J Cell Biol*, 2011. **192**(3): p. 373-82.
206. Amintas, S., et al. *Int J Mol Sci*, 2020. **21**(7).
207. Komine, K., et al. *Anticancer Res*, 2014. **34**(6): p. 3165-8.
208. Zhao, P., et al. *J Cancer*, 2019. **10**(24): p. 6095-6104.
209. Drapkin, B.J., et al. *Cancer Discov*, 2018. **8**(5): p. 600-615.
210. Yu, M., et al. *Science*, 2014. **345**(6193): p. 216-20.
211. Haber, D.A. and V.E. Velculescu. *Cancer Discov*, 2014. **4**(6): p. 650-61.
212. Laprovitera, N., et al. *Mol Oncol*, 2021.
213. Ferracin, M. and M. Negrini. *Expert Review of Molecular Diagnostics*, 2015. **15**(10): p. 1369-1381.
214. Ferracin, M., A. Veronese, and M. Negrini. *Expert Rev Mol Diagn*, 2010. **10**(3): p. 297-308.
215. Laprovitera, N., et al. *Front Oncol*, 2018. **8**: p. 447.
216. Ferracin, M., et al. *Oncotarget*, 2015. **6**(16): p. 14545-55.
217. Ferracin, M., et al. *Journal of Visualized Experiments*, 2016. **2016**(112).
218. Tibshirani, R., et al. *Proc Natl Acad Sci U S A*, 2002. **99**(10): p. 6567-72.
219. Tibshirani, R. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1996. **58**(1): p. 267-288.
220. Barber, L.J.M., S.;Herman, B.; Arezi, B.; Gerlinger, M., *Application Note: Ultra-sensitive Cancer Liquid Biopsy Analysis with the Agilent SureSelectXT HS Target Enrichment Workflow* 2017, Agilent Technologies.
221. Laprovitera, N., et al. *Front Cell Dev Biol*, 2021. **9**: p. 666156.
222. Karczewski, K.J., et al. *Nature*, 2020. **581**(7809): p. 434-443.
223. Povysil, G., et al. *Hum Mutat*, 2017. **38**(7): p. 889-897.
224. Wang, K., M. Li, and H. Hakonarson. *Nucleic Acids Res*, 2010. **38**(16): p. e164.
225. Vaser, R., et al. *Nat Protoc*, 2016. **11**(1): p. 1-9.
226. Adzhubei, I.A., et al. *Nat Methods*, 2010. **7**(4): p. 248-9.
227. Chun, S. and J.C. Fay. *Genome Res*, 2009. **19**(9): p. 1553-61.
228. Schwarz, J.M., et al. *Nat Methods*, 2014. **11**(4): p. 361-2.
229. Reva, B., Y. Antipin, and C. Sander. *Nucleic Acids Research*, 2011. **39**(17): p. e118-e118.
230. Shihab, H.A., et al. *Hum Mutat*, 2013. **34**(1): p. 57-65.
231. Kircher, M., et al. *Nat Genet*, 2014. **46**(3): p. 310-5.
232. Douville, C., et al. *Hum Mutat*, 2016. **37**(1): p. 28-35.
233. Pentheroudakis, G., V. Golfinopoulos, and N. Pavlidis. *Eur J Cancer*, 2007. **43**(14): p. 2026-36.
234. De Carolis, S., et al. *Front Oncol*, 2019. **9**: p. 860.
235. Otter, S., et al. *Clin Oncol (R Coll Radiol)*, 2019. **31**(2): p. 81-90.
236. Piana, A.F., et al. *Virol J*, 2014. **11**: p. 190.
237. Hindson, C.M., et al. *Nat Methods*, 2013. **10**(10): p. 1003-5.
238. Miotto, E., et al. *Cancer Epidemiol Biomarkers Prev*, 2014. **23**(12): p. 2638-42.
239. Zhang, Y., et al. *J Thorac Oncol*, 2015. **10**(10): p. 1444-50.
240. Cristofanilli, M., et al. *N Engl J Med*, 2004. **351**(8): p. 781-91.
241. Cohen, S.J., et al. *J Clin Oncol*, 2008. **26**(19): p. 3213-21.
242. Matsusaka, S., et al. *Cancer Sci*, 2010. **101**(4): p. 1067-71.
243. Lucci, A., et al. *Lancet Oncol*, 2012. **13**(7): p. 688-95.
244. Akhter, M.Z., et al. *Oncogene*, 2018. **37**(16): p. 2089-2103.
245. Dienstmann, R., et al. *Ann Oncol*, 2014. **25**(3): p. 552-563.

246. Turner, N. and R. Grose. *Nat Rev Cancer*, 2010. **10**(2): p. 116-29.
247. Arai, Y., et al. *Hepatology*, 2014. **59**(4): p. 1427-34.
248. Silverman, I.M., et al. 2019. **37**(15_suppl): p. 4080-4080.
249. Deng, N., et al. *Gut*, 2012. **61**(5): p. 673-84.
250. Seo, S., et al. *Oncotarget*, 2017. **8**(20): p. 33844-33854.
251. Byron, S.A., et al. *PLoS One*, 2012. **7**(2): p. e30801.
252. Chang, J., et al. *PLoS One*, 2014. **9**(8): p. e105524.
253. Nakayama, N., et al. *Cancer*, 2010. **116**(11): p. 2621-34.
254. Huang, X., et al. 2020. **38**(15_suppl): p. 558-558.
255. Trizzino, M., et al. *Cell Rep*, 2018. **23**(13): p. 3933-3945.
256. Kadoch, C., et al. *Nat Genet*, 2013. **45**(6): p. 592-601.
257. Tsai, S., et al. *bioRxiv*, 2020: p. 2020.11.16.384446.
258. Shen, J., et al. *Cancer Discov*, 2015. **5**(7): p. 752-67.
259. Shen, J., et al. *Nat Med*, 2018. **24**(5): p. 556-562.
260. Okamura, R., et al. *J Immunother Cancer*, 2020. **8**(1).
261. Silva, M.J., et al. *Environ Mol Mutagen*, 2005. **46**(2): p. 104-15.
262. Kawamoto, Y., et al. *British Journal of Cancer*, 2012. **107**(2): p. 340-344.
263. Benson, A.B., 3rd, et al. *J Natl Compr Canc Netw*, 2014. **12**(7): p. 1028-59.