



**UNIVERSITÀ  
DI TORINO**

**UNIVERSITA' DEGLI STUDI DI TORINO**

**DIPARTIMENTO DI SCIENZE MEDICHE**

**DOTTORATO DI RICERCA IN SCIENZE BIOMEDICHE ED ONCOLOGIA**

**XXXV CICLO**

**An Innovative Experimental Approach to the World of  
"Cancer of Unknown Primary": from Diagnosis to  
Targetable Regulation and Proliferation Mechanisms**

**TESI PRESENTATA DA:  
Dott. ELIANO CASCARDI**

**TUTORS:  
Prof.ssa A. SAPINO  
Prof.ssa C. MARCHIO'**

**COORDINATORE DEL DOTTORATO:  
Prof.re E. HIRSCH**

**ANNI ACCADEMICI: 2019-2023**

**SETTORE SCIENTIFICO-DISCIPLINARE DI AFFERENZA:  
ANATOMIA PATOLOGICA (MED/08)**

# **INDEX**

## **1 Introduction**

- 1.1 Clinical Data
- 1.2 Etiopathogenesis
- 1.3 Genetic Alterations

## **2 Purpose of the Research**

## **3 Results**

- 3.1 AIM 1
  - 3.1.1 Summary and Discussion of the results
- 3.2 AIM 2
  - 3.2.1 Summary and Discussion of the results
- 3.3 AIM 3
  - 3.3.1 Summary and Discussion of the results
- 3.4 Ongoing analysis of PD-L1 as marker of immunotherapy

## **4 Conclusions**

## **References**

### **Attachment 1**

### **Attachment 2**

### **Attachment 3**



# 1 Introduction

## 1.1 Clinical Data

Most cancer patients are diagnosed with a clinically known primary site and characteristic signs and symptoms of that lesion. However, approximately 10-15% of patients show metastatic disease at presentation, and in approximately one-third of these cases, the primary site of disease cannot be determined. [1,2]. These carcinomas are referred to as "Cancer of Unknown Primary" (CUP) [3,4] and are characterized by: (i) early metastatic spread, (ii) unpredictable metastatic patterns, and (iii) lack of specific tissue differentiation markers [4]. A more detailed and accurate definition, proposed by the Royal College of Pathologists [5] and included in the "NICE" (National Institute for Health Care Excellence) Guidelines [6], distinguishes three entities: (1) *Malignancy of Undefined primary Origin* (MUO), defined as metastatic cancer identified by clinical examination or imaging without a known primary site, (2) *Provisional CUP*, defined as metastatic malignancy of epithelial or neuroendocrine origin, diagnosed by cytohistology, without an identified primary site after initial evaluation and multidisciplinary review by specialists, and (3) *CUP* With the same cytohistologic characteristics as preliminary CUP, but for which the site of origin cannot be identified despite thorough examination and multidisciplinary review by specialists.

The most recent guidelines of the Italian Association of Medical Oncology refer to these tumors as "Occult Primary Epithelial Tumors" TEPO" [7]. Non-epithelial histotype neoplasms (e.g., melanoma, sarcoma, lymphoma, germ cell tumors) are not included in CUP by current guidelines.

The first data regarding the incidence of this occult carcinomas date back to 1970,

with Holmes and Foutes reporting that CUP represented 3% of a total of 21,000 carcinoma diagnoses in a 25-year case series [8]. Subsequent studies have generally confirmed this data, with estimates tending to decrease [9-15]. This can be primarily explained by (i) innovation in radiologic technology, (ii) introduction of genetic profiling, and (iii) development of increasingly refined immunohistochemical markers that allow greater diagnostic accuracy in identifying the profile of origin of metastatic diseases. Nonetheless, 80% of CUP patients have a poor prognosis and a survival of only a few months, especially if they present with multiple visceral or bone metastases [16,17]. Conversely, 20% of patients with single disease localizations, superficial lymph node sites, or well-differentiated histotypes have a better prognosis and a median survival of approximately 1-2 years.

## 1.2 Etiopathogenesis

Several theories have been proposed over the years to explain the etiopathogenesis of CUP. In an initial scenario, it is hypothesized that metastatic cells can spread from the primary tumor to distant sites in the early stages and cause an alteration in the microenvironment, resulting in metastasis formation before the primary tumor is identifiable or even before malignant transformation occurs at the site of origin [18]. This theory is supported by the fact that neoplastic cells have different genetic alterations at the site of origin compared to the site of metastasis. [19].

Another scenario concerns the possible presence of tumor microenvironment factors that promote tumor growth at the metastatic site while blocking the growth of these cells (genetically identical to those at the site of origin) in the primary site, resulting in

metastasis without the parallel development of the primary neoplasm [20]. This theory is supported by evidence of clonal relationships between cells at both sites [21].

A final scenario suggests that these neoplasms may originate from undifferentiated pluripotent stem cells capable of differentiating into multiple cell lines [22]. It has been observed that multipotent stem cells are present in connective tissues during adulthood, capable of differentiating into any cell line in the event of neoplastic transformation. According to this hypothesis, CUP can be considered "tumors of adult stem cells" [22].

The current understanding of CUP spread involves the consecutive sequence of clonal proliferation, invasion, and intravasation of tumor cells from the primary tumor, circulation, extravasation into various organs, and colonization at metastatic sites. Once in circulation, cancer cells must survive immune pressures, exit circulation via extravasation, and successfully seed within metastatic sites [23].

### 1.3 Genetic Alterations

Regarding alterations in oncogenes and tumor suppressor genes, it is known that TP53 is the first tumor suppressor gene to be identified in CUP. TP53 mutations are one of the most common genetic alterations in human cancers, with nearly 50% of epithelial malignancies exhibiting biallelic mutations [24].

In CUP, TP53 expression is observed in 70% of cases, with overexpression in 50%. However, TP53 hotspot mutations in exons 5-9 are found in only 26% of CUP cases [25]. This discrepancy between protein expression and gene mutation rate can be attributed in part to the different specificities of antibodies for mutant and wild-type protein and in part to the possible presence of mutations outside the exon 5-9 region. [26].

Activating mutations in RPS7 and RPL11, two genes encoding ribosomal proteins involved in the MDM2/p53 signaling pathway, have also been identified. These genes are implicated in epithelial-mesenchymal transition and increased metastatic potential [27].

Approximately 70% of CUP patients show enrichment in transcripts of proteins involved in DNA damage and homologous recombination repair networks, such as BRCA1, ATM, and CHEK2 (checkpoint kinase 2), suggesting chromosomal instability in CUP [23].

Epidermal Growth Factor Receptor (EGFR) expression is reported in 74-75% of CUP cases, and its overexpression ranges from 4% to 61%. The presence of this receptor correlates with sensitivity to cisplatin-based chemotherapy regimens [28,29].

The expression of tropomyosin-related kinase (NTRK1-related kinases), proteins in the RAS signaling pathway (RAS, RAF, MAPK) have also been observed in a percentage of 6% and 15% CUP cases [30,31].

Expression of c-Kit and PDGFRs has also been reported, although without therapeutic utility or significant prognostic variation. c-Kit expression was reported in 11% and 81%, PDGFRa in approximately 50%, and PDGFRb in approximately 25% of CUP cases. [32,33].

c-Myc is another factor reported to be involved in cell proliferation in CUP. It is expressed in 96% of CUP cases, hyperexpressed in 23%, and amplified in 12%. However, its clinical implications remain unstudied. [34].

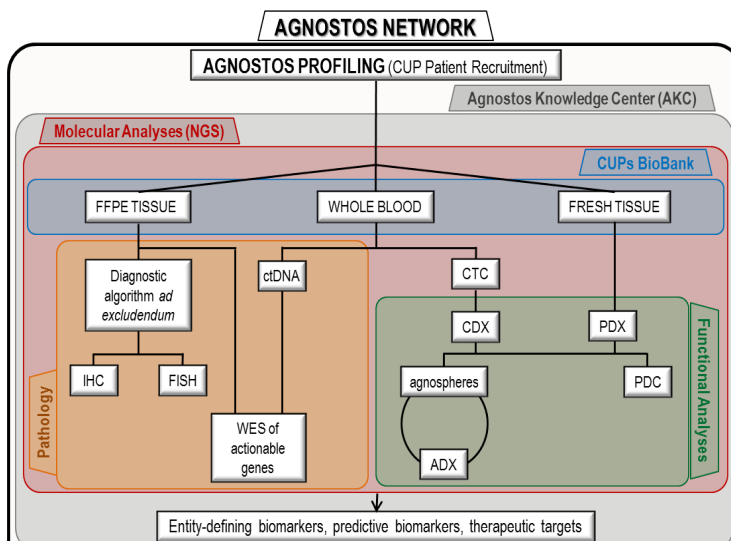
c-Met, a transmembrane receptor binding HGF, shows hyperexpression and somatic mutations associated with invasive growth in CUP. In a cohort of 23 cases of CUP, 30% of them exhibited a c-Met mutation, a significantly higher percentage compared to known primary site tumors, which stands at 4%. These mutated receptors were

functional, suggesting that activating mutations of the MET gene could serve as genetic markers associated with CUP [35].

## 2 Purpose of the Research

Although CUP accounts for only 3-5% of all human tumors, they remain an unsolved medical and diagnostic issue. In particular, the lack of immunohistochemical tissue differentiation markers currently leads to a diagnostic approach of exclusion, following a "what it is not" algorithm. This involves the evaluation of multiple markers of tumor origin, often resulting in excessive sample processing that limits the amount of diagnostic tissue that can be used for molecular analysis, especially in the case of biopsies. Last but not least, this diagnostic workflow can also lead to delayed diagnosis and consequently to a delayed start of treatment for the patient. For all these reasons CUPs epitomize a quintessential metastatic disease, whose understanding can shed light on the largely obscure genetic, molecular and biological principles governing cell dissemination and homing throughout the organism in all cancer subtypes.

Taking all this into account, in the present project we have established a robust, multidisciplinary, and integrated translational platform, as shown in Figure 1, to study the diagnosis, biology, progression and treatment of CUP.



**Figure 1. The AGNOSTOS platform.** From patients recruited through the Agnostos (Greek αγνοστος, meaning 'unknown') profiling protocol, shared by the Agnostos Network of oncological centers, tissues and blood are recovered, biobanked and used for direct diagnostic assessment or derivation of *in vivo* and *in vitro* models for functional analysis, including patient-derived xenografts and agnospheres (CUP cancer stem cells). Molecular analysis, based on deep genomic characterization, pervades all research stages to define diagnostic and predictive biomarkers, and therapeutic targets to inform clinical trials. ADX: agnosphere-derived xenografts, or spheropatient; CDX: circulating tumor cells-derived xenografts; ctDNA: circulating tumor DNA; CTC: circulating tumor cells; FFPE: formalin-fixed paraffin embedded; FISH: fluorescence in situ hybridization; IHC: immunohistochemistry; PDC: patient-derived primary cultures; PDX: patient-derived xenografts; NGS: next generation sequencing.

1. The primary objective of this research project was to develop a simplified and efficient processing of tissue sample to correctly address and standardize the diagnosis of malignancy of "*undefined primary origin*" (MUO).
2. Using biological material of CUP patients, the other goals were (i) to deeply characterize the genomic and transcriptomic landscape of CUPs to discover case-specific and common genetic dependencies and functional vulnerabilities; (ii) to identify the molecular basis that underpins the undifferentiated, stem-like, and invasive features shared by CUPs; (iii) to evaluate potential "drivers" emerging from genetic and functional annotations as targets for personalized medicine.
3. As final aim, still ongoing, we are investigating the expression of PDL-1 in CUP as a marker of immunotherapy.

## 3 Results

### 3.1 AIM 1

[Attachment 1]. **Real-world histopathological approach to malignancy of undefined primary origin (MUO) to diagnose cancers of unknown primary (CUPs).**

Pisacane A#, Cascardi E#, (#contributed equally) et al. *Virchows Arch.* 2023 Mar;482(3):463-475. PMID: 36346458.

#### 3.1.1 Summary and Discussion of the results

The systematic review of the literature [36-40] and the guidelines [5,41,42] regarding MUO are constantly improving, with the aim of defining a diagnostic procedure capable of approximating the tissue of origin as much as possible to better guide therapeutic decisions.

*A) Using data collected from a “real world” experience, we generated a simplified pathology workflow to rule out early metastatic cancer and designed a CUP classification of clinical impact in patients referred to our institute for MUO.*

First, the histologic analysis of 64 MUO referred to our institute, either for initial diagnosis or for second opinion, leads to define two groups of MUO, those easily classified as “carcinomas” and those who need immunohistochemical test because “undifferentiated tumors”. Once carcinomas have been identified, the final step recommended by the guidelines [6] is to determine the probable tissue of origin. Bellizzi [43] reported that through the application of next-generation immunohistochemistry, pathologists can provide increasingly precise answers. The author highlighted the role of tumor morphology in guiding the diagnostic flow for CUPs and further recognized that the metastatic site helps track the primary lesion [43].

We showed that adenocarcinoma was the predominant histotype among



MUO cases classified as “carcinoma” at microscopic analysis. Of note, all the so-called “early metastatic carcinoma” were in the setting of adenocarcinoma and 11% were identifiable by standard histologic features pathognomonic for the tumor of origin (e.g., follicular structures for thyroid cancer, signet ring cells for lobular breast or gastric cancer, etc.). In these cases, the use of a few tissue-specific markers confirmed the organ of origin, thus avoiding specimen waste and guiding imaging studies. Following the recommendation of Bellizzi [43] if the adenocarcinoma morphology was not typical of any tissue of origin, we considered the results of anti-CK7/CK20 antibodies, the patient sex and the site of metastasis to address the additional tissue-specific markers. In some cases, this workflow has resulted in representing a “putative immunophenotype” of the originating tissue. The site of metastases is an useful parameter to be considered in MUO diagnosis and when the lymph nodes are the exclusive site of metastases, it is essential to consider the organs that drain those lymph nodes as the possible site of origin of the tumor [38].

Our work up, which considers the histotype, the site of metastases and the gender, identified another 14% of early metastatic cancers. For example, in women, in the set of CK7+/CK20- MUO adenocarcinoma, estrogen receptor (ER) expression and axillary metastases should be considered to exclude early metastatic breast carcinoma. Using this workflow, we identified 9% of early metastatic breast cancer among MUO. In agreement with these results in the CUPISCO trial, 7.3% were not compatible with CUP because of proof or strong evidence of a breast primary [18]. In the same study the expression of PAX8/WT1 was considered to investigate a possible gynecological origin [18]. We use the same markers to suggest a “putative gynecological immunophenotype” and exclude an early metastatic cancer.

TTF1 and napsin-A are the key lung cancer markers, in CK7+/CK20- adenocarcinomas in men and in ER- biopsies in women. Based on the experience of

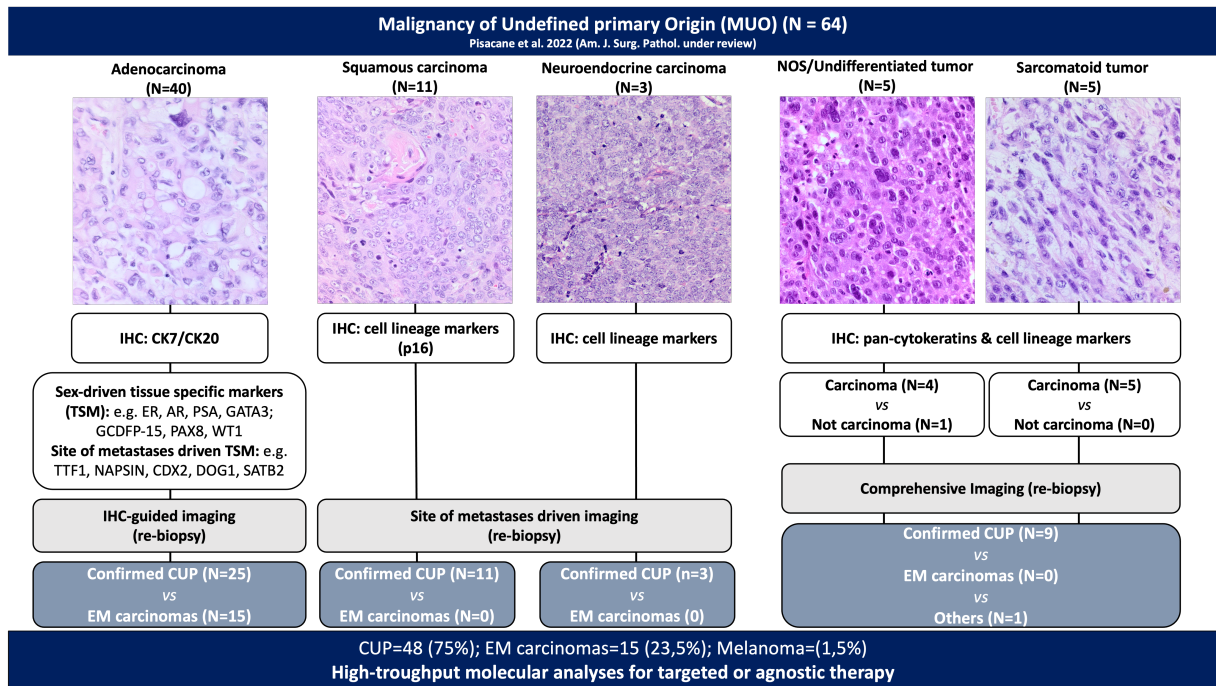
the CUPISCO study [18], all TTF1/napsin A positive cases should be referred to radiologists to rule out lung primitivity. In our series, the lung immunophenotype identified 60% of early metastatic lung cancers.

Finally, in 25% of the adenocarcinoma cohort, the immunophenotyping did not suggest any tissue of origin (adenocarcinoma of undefined immunophenotype) nor any instrumental examination ratified a primary tumor, directly leading to the diagnosis of confirmed CUP.

In cases of CUP with squamous or neuroendocrine morphology, the use of cell differentiation markers is advised to avoid misinterpretation, especially when tumor morphology is heterogenous or poorly differentiated, because these histotypes are considered “favorable CUPs” [44]

In our study, in 15.6% of cases, the pathologist had difficulty defining MUO as carcinoma using standard histologic parameters and had to use a larger panel of immunohistochemistry to differentiate them from sarcoma or melanoma. These “undifferentiated/sarcomatoid CUP” had rapid disease progression. In the CUPISCO trial, one of the reasons for failure was also that 10% of the cases included under “unfavorable CUP” turned out to be non-carcinoma malignancies at central review [18].

Fig. 2 (below) shows the results of the characterization of the MUO case series and the rationale of the workflow starting from histology.



In our case series only 78.7% of formalin-fixed and embedded residual tissue after IHC study was adequate for molecular analyses. Actually, one reason of Cupisco trial failure was the inadequacy of residual tissue, after histopathological diagnosis, for high throughput molecular analyses used to identify possible targets of gene therapy in patients with “unfavorable CUP” [39]. Our cases were tested for genetic analysis using ONCOCARTA that includes known druggable hotspot mutations of ABL1, AKT1, AKT2, BRAF, CDK4, EGFR, ERBB2, FGFR1, FGFR3, FLT3, HRAS, JAK2, KIT, KRAS, MET, NRAS, PDGFA, PIK3CA, and RET genes. 36% of CUPs showed at least one mutation. KRAS and PI3KCA were the most frequently mutated genes and accounted for 75% of all the identified alterations. No specific correlation was found between specific mutations and immunophenotype of adenocarcinomas, sex, histological grade, or survival status.

**B) We showed a clinical impact of our CUP diagnostic workup**

Overall, 42.5% (47/67 of MUO) of patients with confirmed CUP in the present

case series had only lymph node metastases at initial diagnosis, while 57.5% had metastases to other sites (visceral and/or bone and/or muscle and/or brain) with or without lymph node involvement. It is known that CUP-patients with only lymph node metastases have better overall survival [24]. Therefore, to study the clinical impact of our workflow, we focused on patients with confirmed CUP and evaluated the site of first metastases, the site of progression and their possible correlation with the histotype and, in the case of adenocarcinoma, with the putative immunophenotype.

The squamous carcinoma histotype and the “gynecological immunophenotype” (within adenocarcinomas) often presented lymph nodes as the only sites of metastasis. Of note, the NICE guidelines recommend considering metastatic squamous cell carcinoma of the cervical lymph nodes as of head/neck origin and that of the inguinal lymph nodes as of anal/gynecological/lower urological origin [42].

On the opposite, none of the patients of the "biliary-pancreatic immunophenotype" and "pulmonary immunophenotype" adenocarcinomas had only lymph node involvement.

Patients with adenocarcinoma of “undefined immunophenotype,” undifferentiated carcinomas, and sarcomatoid carcinomas had rapid disease progression, regardless of the anatomical location of the cancer at the time of diagnosis and could be considered CUP with “unfavourable prognosis”.

Although we were not able to collect the type of treatment of all patients, we demonstrated that the time to organ progression or death from disease was 6.4 months longer in patients with only lymph node metastases, and that patients had a significantly longer overall survival ( $p < 0.001$ ) regardless of the number of lymph nodes involved and regional location (superficial vs deep and supra vs subdiaphragmatic). These results support lymphadenectomy as a crucial treatment

of these subset of patient. In fact, Wach MM et al. [45] showed that nearly two-thirds of patients undergoing axillary or inguinal lymphadenectomy for metastatic squamous CUP were alive 5 years following lymphadenectomy. In addition, Pouyiourou M et al. advocate to pursue localized treatment with surgery and/ or radiotherapy in single-site and oligometastatic CUP [24].

We acknowledge that our study has some limitations. The cohort was relatively small, and a validation cohort was not available; nevertheless, MUOs are rare entities and a careful clinic-pathological examination of these patients is of utmost importance in order not to misinterpret cases of early metastatic disease as CUPs.

## 3.2 AIM 2

[Attachment 2] **Cancer of unknown primary stem-like cells model multi-organ metastasis and unveil liability to MEK inhibition.** F. Verginelli, A. Pisacane, G. Gambardella, A. D'Ambrosio, E. Candiello, M. Ferrio, M. Panero, L. Casorzo, S. Benvenuti, E. Cascardi, et al. *Nat Commun*, 2021 May 3;12(1):2498. PMID: 33941777

### 3.2.1 Summary and Discussion of the results

A) *While we were defining the critical steps in the diagnosis of CUP, we worked to create in vivo and in vitro models useful to study the CUP biology.*

Fresh human specimens of CUP (3 adenocarcinoma, 1 squamous, 1 sarcomatoid and 1 neuroendocrine carcinoma and the metastatic melanoma as control) were transplanted into immunocompromised NOD/SCID mice (patient-derived xenografts, PDX), resulting in 75% engraftment. Tumor tissue samples derived directly from CUP patients or from CUP-PDX were minced, digested, and seeded to obtain "tumorsphere" and spheroid cultures of stem-like cells called "agnospheres", which reproduced the histological features of the original carcinomas. Agnospheres appeared a few days after seeding and stabilized after an average of 3-4 months, with a 75% of successful rate. Of note, CUP samples that did not engraft belonged to patient with longer survival.

We then performed the genetic characterization of agnospheres and of the corresponding available human tissues, by whole-exome sequencing. Agnospheres display the same genetic makeup of original tumors.

The most interesting observation was that all agnospheres were capable of self-renewing and of long-term propagating at clonal density in the complete absence of any exogenous growth factor, including EGF or FGF, usually required for in vitro isolation and long-term maintenance of the stem phenotype in cancer stem cells from highly aggressive tumors [46]. Independently of exogenous growth factors, all agnospheres displayed the ability to self-sustain the expression of high levels of

transcription factors (TFs) known to be major regulators of self-renewal and stem identity in normal and neoplastic tissues, such as Polycomb repressors [18,46] and reprogramming TFs [19,20,46]. Overall, these data indicate that agnospheres are enriched in cells with transcriptional traits typical of embryonal stem cells, mirrored by functional clonogenic properties, which are fully self-sustained in the absence of exogenous growth factors, thus behaving like “tumoroids” [29]. Such transcriptional features are present also in CUP tissues, attesting to the faithful phenotypic correspondence between the original tumors and the agnospheres, and indicating that CUP tissues are highly enriched in stem-like cells as well. In addition, in agnospheres and original CUP tumors, high and constitutive expression of MYC family genes (c-MYC and/or N-MYC) occurs in the absence of MYC genetic alterations. c-MYC is expressed in the agnospheres derived from the most aggressive cases, diagnosed as “undifferentiated carcinomas” (see attachment 1). N-MYC alone was highly expressed in the single neuroendocrine CUP examined, consistently with its specific role in tissues of neuroectodermal origin [47].

We also evaluated the analogies between original CUP and the *in vivo* PDX models obtained from subcutaneous transplantation of agnospheres. Thus, we compared CUP immunohistochemical pattern and clinical progression with those of PDX. To study tumor progression PDX were euthanized 10 days after subcutaneous injection, and explanted organs were studied by *ex vivo* organ luminescence and were also processed to obtain FFPE samples to evaluate the histology and the number of metastases, if present. These are the main results:

- (i) The immunohistochemical pattern of original CUP was maintained in PDX tumor.
- (ii) Multiple metastases were detected in the same site/organs of metastases observed in patients. Furthermore, dissemination occurred well before the tumor

became palpable, suggesting that cells can initiate the process without prior local expansion and, probably, without the induction of a pre-metastatic niche by an already established tumor [18]. Such a rapid lymphogenous and hematogenous widespread metastatization is rather uncommon when organoids obtained from other aggressive carcinomas are transplanted subcutaneously. Overall, these findings attest that agnospheres, in particular those from aggressive CUPs, retain a high spontaneous and multi-organ metastatic ability, which recapitulates the CUP.

(iii) In PDX transplanted with aggressive agnospheres, metastatic burden was lethal within 4 months, and also in this case postmortem analysis showed multiple metastatic sites in approximately 75% of mice.

*B) We demonstrated that agnospheres and CUP-PDX were sensible to MEK1/2 inhibitor.*

Agnospheres display high levels of constitutively phosphorylated MEK1/2. We therefore reasoned that the multiple mechanisms sustaining agnosphere autonomous proliferation should converge onto the MAPK pathway. CUP-PDX obtained from three representative agnospheres (2 carcinoma and 1 neuroendocrine carcinoma) were challenged with a selective, clinically approved, MEK1/2 inhibitor, trametinib, used in BRAF-mutated melanomas [43,46,48]. In the two PDX (derived from carcinoma), trametinib was effective and induced dramatic growth arrest within 4 to 10 days, accompanied by long-term inhibition of ERK1/2 phosphorylation, the direct target of MEK1/2, and downregulation of c-MYC, known to be tightly regulated by proliferative signals via the MAPK pathway [42,49]. This was accompanied by (i) proportional downregulation of cMYC transcriptional target Cyclin D1, responsible for G1–S cell cycle progression via CDK4/6-mediated pRB hyperphosphorylation, which



was consistently reduced as well; (ii) stabilization of the cell cycle inhibitor p27-KIP1, known to be prevented by c-MYC.

Histological analysis of Trametinib treated PDX tumors showed ample central necrosis and a lower number and size of metastatic vascular emboli as compared with controls.

C) *We elaborated a “trametinib response signature” able to predict sensitivity to trametinib.*

To this aim, we correlated publicly available data on the trametinib response of 445 cancer cell lines (CCLs) with the respective gene expression profiles [36,37]. A specific “trametinib response signature” was identified. Agnospheres originated from different tumors were then tested for this signature, which correctly anticipated the response to trametinib and it was retrieved also in the matched patients’ tissues. Finally, we analysed different synchronous metastases collected at warm autopsy from a single patient (AGN43) [50], showing that all shared the “trametinib sensitivity signature”. This finding adds to the previously shown genetic homogeneity of these metastases [50] and suggests that, at least in some cases, the cell-autonomous behavior of CUP stem- like cells may translate in remarkable biological homogeneity and drug sensitivity across metastatic sites, with favorable implications for CUP therapy.

### 3.3 AIM 3

[Attachment 3] **Mutated axon guidance gene PLXNB2 sustains growth and invasiveness of stem cells isolated from cancers of unknown primary.** Brundu S, Napolitano V, Franzolin G, Lo Cascio E, Mastrantonio R, Sardo G, Cascardi E, et al. *EMBO Mol Med.* 2023 Mar 8;15(3):e16104. PMID: 36722641

#### 3.3.1 Summary and Discussion of the results

*A) We have identified the overexpression of the axon guidance gene PLXNB2 in CUP.*

This study was conducted taking advantages of the CUP patient cohort collected in our institution and of the previously obtained in vitro and in vivo models. The data of whole exome sequencing (WES) of metastases in 14 CUP were analysed to investigate whether any functional pathways were more frequently affected by genetic changes. We found three subgroups: five tumors could be defined as hypomutated (showing < 50 genetic alterations per exome), six were normomutated (showing 50 to 250 mutations), and three were hypermutated (having more than 106 mutations).

We comparatively analysed the status of genes clustered in 186 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and compared to datasets derived from 33 different cohorts of patients with other tumor types from TCGA. Based on 24 informative comparisons between tumors, some pathways were found to be more frequently mutated in CUPs, among these the Axon Guidance pathway, which is known for its emerging role in cancer progression and metastasis, beyond to the fact that this pathway included PLXNA4, one of the most frequently mutated genes in our CUP cohort.

To address the potential functional involvement of these mutated genes in the CUP phenotype, we integrated WES analysis with gene expression profiling. We identified three genes, of the Axon Guidance genes, highly expressed in all CUP:

CXCR4, PPP3CA and PLXNB2.

A PLXNB2 mutation found in our CUP cohort, and not previously reported in COSMIC, encoding the amino acid change p.G842C, was particularly interesting based on our in silico structural analysis. Of note, this change was preserved in 14 independent metastatic lesions biopsied in the CUP patient AGN43.

We then decided to focus our study on the potential pathogenic role of this mutant axon guidance receptor in CUP cells.

*B. We demonstrated a new pathogenic role of G842C mutant axon guidance receptor in CUP cells.*

*1B)* The G842C mutation of PlxnB2 replaced a conserved residue in the extracellular domain, with expected impact on receptor conformation. To predict the potential detrimental impact of the G842C amino acid change in PlxnB2, the structure of this protein was retrieved from the AlphaFold database and three-dimensional model structures of wild-type human plexin and the G842C variant. Molecular dynamics simulations were then performed to test the system. We noted that the G842C mutation results in a change in the native domain that causes destabilization of the tertiary structure of the IPT3 domain.

G842C-PlxnB2 basically acts as a putative gain-of-function mutation, leading to ligand-independent receptor activation in overexpressing cells. Knockdown of G842C-mutated, but not wild-type, PlxnB2 impairs CUP cell viability, clonogenic capacity, and tumorigenic growth in mice.

*2B)* To assess the functional relevance of the endogenous G842C-mutated PlxnB2 protein, we knocked down its expression in agnospheres derived from CUP patient AGN43, presenting with multiple metastases and evaluated cell proliferation and stem cell frequency. For comparison, we performed similar experiments in

agnospheres derived from patients AGN901, AGN906, and AGN67, which express wild-type PlxnB2. Knockdown of PlxnB2 had no significant impact on the growth of agnospheres carrying wild-type PlxnB2; in contrast, depletion of G842C-mutated PlxnB2 in AS43 resulted in marked inhibition of cell viability and growth compared to respective controls. In addition, G842C-PlxnB2 could impact tumor cell self-renewal and tumorigenesis as shown by experiments using limiting dilution assays (LDA) and gene silencing.

Based on the above findings, to better investigate the role of mutated PlxnB2 in tumor progression, we generated agnosphere-derived from NOD-SCID mice tumor obtained from AGN43 or AGN901 with PlxnB2-depleted and monitored tumor growth by periodic calibration. Our results indicated that expression of G842C mutated PlxnB2 is specifically required for the growth of AGN43-derived tumors. Indeed, while we observed a marked reduction in tumor volume in PlxnB2-silenced PDX43 mice, there was no significant difference between silenced and control tumors in PDX901 bearing wild-type PlxnB2, further suggesting that the G842C mutation promotes CUP growth in vivo.

3B) We next aimed to identify the signalling mechanisms potentially responsible for the promotion of CUP cell stemness, proliferation, and invasiveness by G842C-mutated PlxnB2. Ligand-dependent PlexinB2 activity has previously been shown to stimulate ErbB2 tyrosine phosphorylation and oncogenic signaling in breast cancer cells [51]. In the previous study of our group [attachment 2] [50], we showed that the EGFR tyrosine kinase is basally phosphorylated at low levels in CUP cells carrying a wild-type PlxnB2 (AGN901, AGN906), whereas it is highly phosphorylated in AGN43 cells, carrying the modified G842C receptor. EGFR is an important oncogenic promoter in human tumors, particularly in cases of gene amplification and overexpression; however, AGN43 does not carry EGFR mutations or elevated

expression, so we hypothesized that the presence of G842C-PlxnB2 might promote its non-canonical activation.

First, we tested the impact of silencing mutated PlexinB2 in CUP AGN43 cells and found that this resulted in a net decrease of EGFR tyrosine phosphorylation, confirming that this altered axon guidance signal controls an important oncogenic pathway in CUP cells. We then analyzed phospho-EGFR levels in basally unmutated CUP cells (AGN906 and AGN901), after transduction of the wild-type or mutated PlxnB2 isoform with G842C. Mutated plexin elicited significantly higher EGFR phosphorylation, compared to controls. This indicates that G842C-PlxnB2 expression is necessary and sufficient to support EGFR phosphorylation in CUP-derived agnospheres. However, we were unable to obtain conclusive evidence on a quantitative difference between the association of EGFR with mutated versus wild-type PlxnB2. Our data suggest that the two receptors might interact in a dynamic complex on the cell surface, which could enable trans-regulation of EGFR through conformational changes driven by mutation in PlxnB2. Consistent with these findings, cetuximab an inhibitor of EGFR signalling, produced inhibition of AGN43 cell viability and clonogenic capacity, indicating that these cells are functionally dependent, in contrast to wild-type PlxnB2 agnospheres. Thus, increased EGFR activity driven by genetically altered PlxnB2 controls intracellular signalling and proliferative self-sufficiency in CUP cells.

We finally investigated whether EGFR kinase activity might be responsible for the increased invasiveness of CUP cells driven by overexpression of G842C-mutated PlxnB2. The high invasiveness of CUP cells elicited by the PlexinB2 G842C mutant isoform was almost completely abrogated by the EGFR inhibitor, indicating that the PlexinB2 axon guidance receptor is functionally positioned upstream of the EGFR kinase in the regulation of CUP cells.

### 3.4 Ongoing analysis of PDL1 as marker of immunotherapy

CUP treatment is still empirical, and immunotherapy could be an option. The standard marker used to enroll patient for immunotherapy is PDL1, analyzed by immunohistochemistry [52]. In one study, using the Tumor Proportion Score (TPS), defined as the number of positive tumor cells divided by the total number of viable tumor cells multiplied by 100%, and a cutoff of >5% of PDL1 moderate (2+) membranous positive expression in cancer cells, 22.5% of CUPs resulted to be PDL1 positive [28]. Using TPS and the same cut off, 25/29 (86%) CUP patients were eligible to pembrolizumab treatment in a single phase 2 clinical trial. The study showed encouraging clinical activity with an acceptable safety profile in patients with previously treated CUP [53].

There are currently multiple different PD-L1 IHC assays that have been developed as companion or complementary diagnostic assays for various immunotherapy agents, each with varying antibody clones, IHC detection systems, scoring cutoffs, testing of tumor cells vs. tumor-infiltrating immune cells, and performance metrics. This has led to ongoing uncertainties regarding the ability to interchange and cross-compare results between different testing platforms [54]

Taking advantage of the first objective of standardizing the workflow of CUP diagnosis, we thought it would be useful to standardize the assessment of PD-L1 in our CUP series using the PD-L1 antibody (Agilent clone 22C3) from the pharmDx kit on Dako's Automated Link 48 platform. We stained the 15 out of 47 CUP cases whose samples were still suitable for immunohistochemical analysis after the applying our workflow. To assess PDL1 expression, three different pathologists applied the Combined Positive Score (CPS), which evaluates the number of cells (tumor cells, lymphocytes, and macrophages) positive for anti-PD-L1 antibody among all cells present and multiplies the result by 100. A CPS score  $\geq 1$ , considered

positive, was reported independently by the three raters on the same seven cases. Three of these cases were adenocarcinomas, 2 were squamous cell carcinomas and 2 were sarcomatoid carcinomas(Fig. 3). This preliminary result shows that a high percentage of CUP express PDL1 at high level, reinforcing the possible response to immunotherapy.

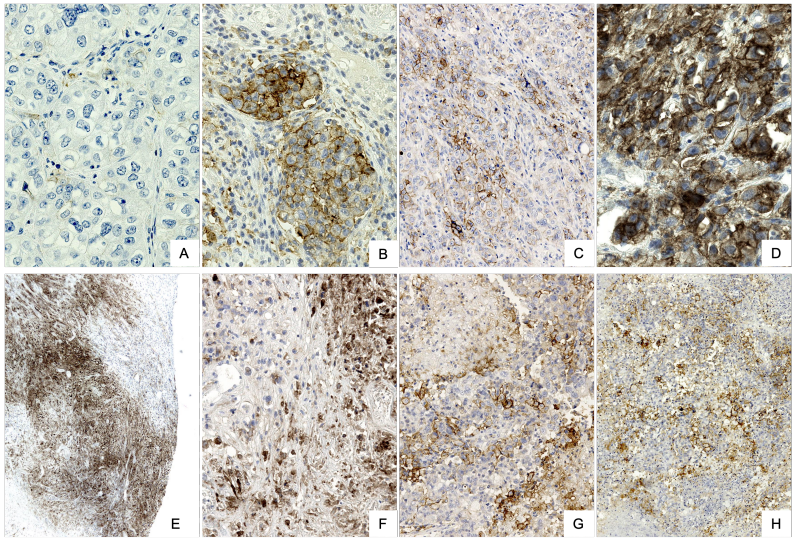


Fig. 3: PDL1- immunohistochemical expression in CUP adenocarcinomas (A: negative B, C, D positive); sarcomatoid CUP (E, F) and squamous CUP (H).

## 4 Conclusions

The discovery of metastatic cancer of unknown origin at the time of initial diagnosis is a dramatic event for patients and a frustrating experience for clinicians due to the lack of specific therapeutic options.

Therefore, we designed a multidisciplinary project to effectively diagnose CUP and to establish reliable *in vitro* and *in vivo* models to study CUP genetic alterations and to identify potential "drivers" as targets for personalized medicine.

First, using data collected from a real-life experience, we envisage a streamlined pathological workup to rule out CUPs in patients presenting with undefined primary origin (MUO) demonstrating that:

- standard histology is of crucial importance and primarily guides the diagnostic process of MUO. We confirm that CUP cases are carcinomas and adenocarcinomas are the most common histotype.
- MUO's immunohistochemical analysis to exclude early metastatic cancer is more effective and reduces tissue wastage when guided by histology and site of metastasis and sex of the patient.
- A "putative immunophenotype", which retains some flavor of the possible organ of origin, of CUP adenocarcinomas influences the progression of the disease.

The Royal Collage of Pathologists stressed the importance of the histotype to address immunohistochemical analyses correctly. The histotypes considered by ESMO guidelines [6] are well- and moderately differentiated adenocarcinomas, squamous cell carcinomas, carcinomas with neuroendocrine differentiation, poorly differentiated carcinomas (including poorly differentiated adenocarcinomas), and undifferentiated neoplasms. The same guidelines [6] indicate CK7 and CK20 as



primary markers in the basic immunohistochemical workup of CUPs, followed by other tissue-specific markers to recognize the tissue of origin. However, recently, Pauli et al. [44] highlighted the challenge in identifying patients with CUP according to ESMO guidelines. We think that our proposed diagnostic algorithm, although limited by the small sample size, can guide the pathologist in the diagnostic process of CUP and provide prognostic evidence in CUP patients.

Other important goals of the project that we achieved were:

- we obtain reproducible *in vitro* models, called "agnospheres", and *in vivo* CUP mouse models derived from "cancer stem cells", which exactly replicate the multi-organ metastatic pattern of CUP and retain the driver genetic alterations of the original tumor.

- The WES analysis carried out did not find genetic changes in known oncogenes or tumor suppressor genes consistently associated with CUP samples, in agreement to previous evidence [9].

- Agnospheres rely on constitutively active MEK signaling, which confers sensitivity to MEK inhibition *in vitro* and *in vivo*. The specific MEK1/2 inhibitor Trametinib induced growth arrest and apoptosis. When, administered *in vivo* in PDX, Trametinib halted the growth of experimental tumors, inducing necrosis and preventing dissemination.

- Such sensitivity is associated with a transcriptomic signature predicting that more than 70% of CUP patients could be eligible to MEK inhibition.

These results shed light on CUP biology and unveil an opportunity for therapeutic intervention with Trametinib.

In an another line of investigation, the axon guidance pathway caught our attention, due to its emerging role in cancer progression and metastasis dissemination.

- In silico analysis highlighted PlexinB2 p.G842C mutation as particularly intriguing. Indeed, wild-type PlexinB2 is implicated in the regulation of cancer cell proliferation, invasiveness, and metastatic spreading [55], but PlexinB2 mutations were never described before.

- The PlexinB2-G842C mutation is conserved in CUP *in vitro* and *in vivo* models.

- The PlexinB2-G842C acts as a gain-of-function mutation promoting PlexinB2 signaling. Silence of PlexinB2-G842C expression results in a significant decrease of stem cell frequency, *in vitro* proliferation and *in vivo* tumorigenicity, while PlexinB2-G842C transduction increases proliferative rate and tumorigenicity indicating that this mutation upregulates programs highly relevant for cancer progression.

- Finally, we demonstrated that PlexinB2-G842C can promote CUP cell stemness, proliferation and invasiveness by sustaining EGFR activation.

The finding of mutations in axon guidance genes, PLXNB2 in particular, may provide novel genetic biomarkers guiding CUP disease management. Moreover, the identification of EGFR signaling deregulation, as functional consequence of the aberrant activation of axon guidance pathways, may prompt the development of novel therapeutic approaches for CUP patients carrying the implicated mutations.

Our project has led to important results in the diagnosis and biology of CUP and shed light on the unexpected genetic alterations underlying the origin of CUP. The possible transfer of these findings to patients is being studied to design specific clinical trials.

## References

1. Oien, K.A. Pathologic evaluation of unknown primary cancer. *Semin Oncol* **2009**, *36*, 8-37, doi:10.1053/j.seminoncol.2008.10.009.
2. NCCN. Occult Primary (Clinical Practice Guidelines in Oncology). **2019**, at <https://www2.tri-kobe.org/nccn/guideline/occult/english/occult.pdf>.
3. Kolling, S.; Ventre, F.; Geuna, E.; Milan, M.; Pisacane, A.; Boccaccio, C.; Sapino, A.; Montemurro, F. "Metastatic Cancer of Unknown Primary" or "Primary Metastatic Cancer"? *Front Oncol* **2019**, *9*, 1546, doi:10.3389/fonc.2019.01546.
4. Rassy, E.; Pavlidis, N. Progress in refining the clinical management of cancer of unknown primary in the molecular era. *Nat Rev Clin Oncol* **2020**, *17*, 541-554, doi:10.1038/s41571-020-0359-1.
5. Schofield, J.B.; Oien, K. G167 Dataset for histopathological reporting of cancer of unknown primary (CUP) and malignancy of unknown primary origin (MUO). **2018**.
6. Cancer, N.C.C.f. Diagnosis and management of metastatic malignant disease of unknown primary origin. National Institute for Health and Clinical Excellence: Guidance; 2010.
7. AIOM. Linee Guida AIOM Tumori Epiteliali Occulti (TEPO). **2020**, at [https://www.aiom.it/wp-content/uploads/2020/12/2020\\_AIOM\\_LG\\_TEPO.pdf](https://www.aiom.it/wp-content/uploads/2020/12/2020_AIOM_LG_TEPO.pdf).
8. Holmes, F.F.; Fouts, T.L. Metastatic cancer of unknown primary site. *Cancer* **1970**, *26*, 816-820, doi:10.1002/1097-0142(197010)26:4<816::aid-cnrcr2820260413>3.0.co;2-r.
9. Altman, E.; Cadman, E. An analysis of 1539 patients with cancer of unknown primary site. *Cancer* **1986**, *57*, 120-124, doi:10.1002/1097-0142(19860101)57:1<120::aid-cnrcr2820570124>3.0.co;2-m.
10. Urban, D.; Rao, A.; Bressel, M.; Lawrence, Y.R.; Mileskin, L. Cancer of unknown primary: a population-based analysis of temporal change and socioeconomic disparities. *Br J Cancer* **2013**, *109*, 1318-1324, doi:10.1038/bjc.2013.386.
11. Brustugun, O.T.; Helland, A. Rapid reduction in the incidence of cancer of unknown primary. A population-based study. *Acta Oncol* **2014**, *53*, 134-137, doi:10.3109/0284186X.2013.783230.
12. Brewster, D.H.; Lang, J.; Bhatti, L.A.; Thomson, C.S.; Oien, K.A. Descriptive epidemiology of cancer of unknown primary site in Scotland, 1961-2010. *Cancer Epidemiol* **2014**, *38*, 227-234, doi:10.1016/j.canep.2014.03.010.
13. Binder, C.; Matthes, K.L.; Korol, D.; Rohrmann, S.; Moch, H. Cancer of unknown primary- Epidemiological trends and relevance of comprehensive genomic profiling. *Cancer Med* **2018**, *7*, 4814-4824, doi:10.1002/cam4.1689.
14. Bevier, M.; Sundquist, J.; Hemminki, K. Incidence of cancer of unknown primary in Sweden: analysis by location of metastasis. *Eur J Cancer Prev* **2012**, *21*, 596-601, doi:10.1097/CEJ.0b013e3283523468.
15. Shu, X.; Sundquist, K.; Sundquist, J.; Hemminki, K. Time trends in incidence, causes of death, and survival of cancer of unknown primary in Sweden. *Eur J Cancer Prev* **2012**, *21*, 281-288, doi:10.1097/CEJ.0b013e32834c9ceb.
16. Losa, F.; Fernandez, I.; Etzaniz, O.; Gimenez, A.; Gomila, P.; Iglesias, L.; Longo, F.; Nogales, E.; Sanchez, A.; Soler, G. SEOM-GECOD clinical guideline for unknown primary cancer (2021). *Clin Transl Oncol* **2022**, *24*, 681-692, doi:10.1007/s12094-022-02806-x.
17. Huey, R.W.; Smaglo, B.G.; Estrella, J.S.; Matamoros, A.; Overman, M.J.; Varadhachary, G.R.; Raghav, K.P.S. Cancer of Unknown Primary Presenting as Bone-Predominant or Lymph Node-Only Disease: A Clinicopathologic Portrait. *Oncologist* **2021**, *26*, e650-e657, doi:10.1002/onco.13700.
18. Lopez-Lazaro, M. The migration ability of stem cells can explain the existence of cancer of unknown primary site. Rethinking metastasis. *Oncoscience* **2015**, *2*, 467-475, doi:10.18632/oncoscience.159.
19. Joosse, S.A.; Pantel, K. Genetic traits for hematogeneous tumor cell dissemination in cancer patients. *Cancer Metastasis Rev* **2016**, *35*, 41-48, doi:10.1007/s10555-016-9611-7.
20. Tarin, D. Clinical and biological implications of the tumor microenvironment. *Cancer Microenviron* **2012**, *5*, 95-112, doi:10.1007/s12307-012-0099-6.
21. Speel, E.J.M.; van de Wouw, A.J.; Claessen, S.M.; Haesevoets, A.; Hopman, A.H.; van der Wurff, A.A.; Osieka, R.; Buettner, R.; Hillen, H.F.; Ramaekers, F.C. Molecular evidence for a clonal relationship between multiple lesions in patients with unknown primary adenocarcinoma.

- International journal of cancer* **2008**, *123*, 1292-1300.
22. Greco, F.A. Cancer of unknown primary site: still an entity, a biological mystery and a metastatic model. *Nat Rev Cancer* **2014**, *14*, 3-4, doi:10.1038/nrc3646.
  23. Rassy, E.; Assi, T.; Pavlidis, N. Exploring the biological hallmarks of cancer of unknown primary: where do we stand today? *Br J Cancer* **2020**, *122*, 1124-1132, doi:10.1038/s41416-019-0723-z.
  24. Soussi, T.; Ishioka, C.; Claustres, M.; Beroud, C. Locus-specific mutation databases: pitfalls and good practice based on the p53 experience. *Nat Rev Cancer* **2006**, *6*, 83-90, doi:10.1038/nrc1783.
  25. Bar-Eli, M.; Abbruzzese, J.L.; Lee-Jackson, D.; Frost, P. p53 gene mutation spectrum in human unknown primary tumors. *Anticancer Res* **1993**, *13*, 1619-1623.
  26. Vousden, K.H.; Vande Woude, G.F. The ins and outs of p53. *Nat Cell Biol* **2000**, *2*, E178-180, doi:10.1038/35036427.
  27. Kurahashi, I.; Fujita, Y.; Arai, T.; Kurata, T.; Koh, Y.; Sakai, K.; Matsumoto, K.; Tanioka, M.; Takeda, K.; Takiguchi, Y.; et al. A microarray-based gene expression analysis to identify diagnostic biomarkers for unknown primary cancer. *PLoS One* **2013**, *8*, e63249, doi:10.1371/journal.pone.0063249.
  28. Gatalica, Z.; Xiu, J.; Swensen, J.; Vranic, S. Comprehensive analysis of cancers of unknown primary for the biomarkers of response to immune checkpoint blockade therapy. *Eur J Cancer* **2018**, *94*, 179-186, doi:10.1016/j.ejca.2018.02.021.
  29. Hainsworth, J.D.; Lenington, W.J.; Greco, F.A. Overexpression of Her-2 in patients with poorly differentiated carcinoma or poorly differentiated adenocarcinoma of unknown primary site. *J Clin Oncol* **2000**, *18*, 632-635, doi:10.1200/JCO.2000.18.3.632.
  30. Mauri, G.; Valtorta, E.; Cerea, G.; Amatu, A.; Schirru, M.; Marrapese, G.; Fiorillo, V.; Recchimuzzo, P.; Cavenago, I.S.; Bonazzina, E.F.; et al. TRKA expression and NTRK1 gene copy number across solid tumours. *J Clin Pathol* **2018**, *71*, 926-931, doi:10.1136/jclinpath-2018-205124.
  31. Pentheroudakis, G.; Kotteas, E.A.; Kotoula, V.; Papadopoulou, K.; Charalambous, E.; Cervantes, A.; Ciuleanu, T.; Fountzilias, G.; Pavlidis, N. Mutational profiling of the RAS, PI3K, MET and b-catenin pathways in cancer of unknown primary: a retrospective study of the Hellenic Cooperative Oncology Group. *Clin Exp Metastasis* **2014**, *31*, 761-769, doi:10.1007/s10585-014-9666-1.
  32. Dova, L.; Pentheroudakis, G.; Golfopoulou, V.; Malamou-Mitsi, V.; Georgiou, I.; Vartholomatos, G.; Ntemou, A.; Fountzilias, G.; Pavlidis, N. Targeting c-KIT, PDGFR in cancer of unknown primary: a screening study for molecular markers of benefit. *J Cancer Res Clin Oncol* **2008**, *134*, 697-704, doi:10.1007/s00432-007-0341-7.
  33. Koo, J.S.; Kim, H. Hypoxia-related protein expression and its clinicopathologic implication in carcinoma of unknown primary. *Tumour Biol* **2011**, *32*, 893-904, doi:10.1007/s13277-011-0190-5.
  34. Clynick, B.; Dessauvage, B.; Sterrett, G.; Harvey, N.T.; Allcock, R.J.N.; Saunders, C.; Erber, W.; Meehan, K. Genetic characterisation of molecular targets in carcinoma of unknown primary. *J Transl Med* **2018**, *16*, 185, doi:10.1186/s12967-018-1564-x.
  35. Stoyianni, A.; Goussia, A.; Pentheroudakis, G.; Siozopoulou, V.; Ioachim, E.; Krikelis, D.; Golfopoulou, V.; Cervantes, A.; Bobos, M.; Fotsis, T.; et al. Immunohistochemical study of the epithelial-mesenchymal transition phenotype in cancer of unknown primary: incidence, correlations and prognostic utility. *Anticancer Res* **2012**, *32*, 1273-1281.
  36. Lin, F.; Liu, H. Immunohistochemistry in undifferentiated neoplasm/tumor of uncertain origin. *Archives of pathology & laboratory medicine* **2014**, *138*, 1583-1610.
  37. Bochtler, T.; Löffler, H.; Krämer, A. Diagnosis and management of metastatic neoplasms with unknown primary. In Proceedings of the Seminars in diagnostic pathology, 2018; pp. 199-206.
  38. Laprovitera, N.; Riefolo, M.; Ambrosini, E.; Klec, C.; Pichler, M.; Ferracin, M. Cancer of unknown primary: challenges and progress in clinical management. *Cancers* **2021**, *13*, 451.
  39. Kato, S.; Alsafar, A.; Walavalkar, V.; Hainsworth, J.; Kurzrock, R. Cancer of Unknown Primary in the Molecular Era. *Trends in cancer* **2021**.
  40. Qaseem, A.; Usman, N.; Jayaraj, J.S.; Janapala, R.N.; Kashif, T. Cancer of unknown primary: a review on clinical guidelines in the development and targeted management of patients with the unknown primary site. *Cureus* **2019**, *11*.
  41. Fizazi, K.; Greco, F.A.; Pavlidis, N.; Daugaard, G.; Oien, K.; Pentheroudakis, G.; Committee, E.G. Cancers of unknown primary site: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **2015**, *26 Suppl 5*, v133-138, doi:10.1093/annonc/mdv305.
  42. National Institute for Health and Care Excellence. Metastatic malignant disease of unknown primary origin in adults: diagnosis and management. [*NICE Guideline No. 104*].

<https://www.nice.org.uk/guidance/cg104> (2019).

43. Bellizzi, A.M. An algorithmic immunohistochemical approach to define tumor type and assign site of origin. *Advances in anatomic pathology* **2020**, *27*, 114.
44. Pauli, C.; Bochtler, T.; Mileshkin, L.; Baciarello, G.; Losa, F.; Ross, J.S.; Pentheroudakis, G.; Zarkavelis, G.; Yalcin, S.; Özgüroğlu, M. A challenging task: identifying patients with cancer of unknown primary (CUP) according to ESMO guidelines: the CUPISCO trial experience. *The oncologist* **2021**, *26*, e769-e779.
45. Wach, M.M.; van Beek, E.; Ayabe, R.; Ruff, S.; Brown, Z.; Goldman, D.A.; Zambirinis, C.P.; Gholami, S.; Pulitzer, M.; Hernandez, J. Metastatic squamous cell carcinoma of known and unknown primary origin treated with axillary or inguinal lymphadenectomy. *The American Journal of Surgery* **2018**, *216*, 963-968.
46. Pastrana, E.; Silva-Vargas, V.; Doetsch, F. Eyes wide open: a critical review of sphere-formation as an assay for stem cells. *Cell stem cell* **2011**, *8*, 486-498.
47. Gilligan, T.; Seidenfeld, J.; Basch, E.; Einhorn, L.; Fancher, T.; Smith, D.; Stephenson, A.; Vaughn, D.; Cosby, R.; Hayes, D. American Society of Clinical Oncology clinical practice guideline on uses of serum tumor markers in adult males with germ cell tumors. *J Clin Oncol* **2010**, *28*, 3388-3404.
48. Shao, Y.; Liu, X.; Hu, S.; Zhang, Y.; Li, W.; Zhou, X.; Wang, Q.; Hou, Y.; Chen, Y.; Wang, Y. Sentinel node theory helps tracking of primary lesions of cancers of unknown primary. *BMC cancer* **2020**, *20*, 1-8.
49. Ross, J.S.; Sokol, E.S.; Moch, H.; Mileshkin, L.; Baciarello, G.; Losa, F.; Beringer, A.; Thomas, M.; Elvin, J.A.; Ngo, N. Comprehensive genomic profiling of carcinoma of unknown primary origin: Retrospective molecular classification considering the CUPISCO study design. *The oncologist* **2021**, *26*, e394-e402.
50. Verginelli, F.; Pisacane, A.; Gambardella, G.; D'Ambrosio, A.; Candiello, E.; Ferrio, M.; Panero, M.; Casorzo, L.; Benvenuti, S.; Cascardi, E. Cancer of unknown primary stem-like cells model multi-organ metastasis and unveil liability to MEK inhibition. *Nature communications* **2021**, *12*, 1-16.
51. Gurrupu, S.; Pupo, E.; Franzolin, G.; Lanzetti, L.; Tamagnone, L. Sema4C/PlexinB2 signaling controls breast cancer cell growth, hormonal dependence and tumorigenic potential. *Cell Death & Differentiation* **2018**, *25*, 1259-1275.
52. Haratani, K.; Hayashi, H.; Takahama, T.; Nakamura, Y.; Tomida, S.; Yoshida, T.; Chiba, Y.; Sawada, T.; Sakai, K.; Fujita, Y.; et al. Clinical and immune profiling for cancer of unknown primary site. *J Immunother Cancer* **2019**, *7*, 251, doi:10.1186/s40425-019-0720-z.
53. Raghav, K.P.; Stephen, B.; Karp, D.D.; Piha-Paul, S.A.; Hong, D.S.; Jain, D.; Onwugaje, D.O.C.; Abonofal, A.; Willett, A.F.; Overman, M. Efficacy of pembrolizumab in patients with advanced cancer of unknown primary (CUP): a phase 2 non-randomized clinical trial. *Journal for immunotherapy of cancer* **2022**, *10*.
54. Hirsch, F.R.; McElhinny, A.; Stanforth, D.; Ranger-Moore, J.; Jansson, M.; Kulangara, K.; Richardson, W.; Towne, P.; Hanks, D.; Vennapusa, B. PD-L1 immunohistochemistry assays for lung cancer: results from phase 1 of the blueprint PD-L1 IHC assay comparison project. *Journal of Thoracic Oncology* **2017**, *12*, 208-222.
55. Brundu, S.; Napolitano, V.; Franzolin, G.; Lo Cascio, E.; Mastrantonio, R.; Sardo, G.; Cascardi, E.; Verginelli, F.; Sarnataro, S.; Gambardella, G. Mutated axon guidance gene PLXNB2 sustains growth and invasiveness of stem cells isolated from cancers of unknown primary. *EMBO Molecular Medicine* **2023**, *15*, e16104.



# Real-world histopathological approach to malignancy of undefined primary origin (MUO) to diagnose cancers of unknown primary (CUPs)

Alberto Pisacane<sup>1</sup> · Eliano Cascardi<sup>1,2</sup> · Enrico Berrino<sup>1,2</sup> · Alessio Polidori<sup>1</sup> · Ivana Sarotto<sup>1</sup> · Laura Casorzo<sup>1</sup> · Mara Panero<sup>1</sup> · Carla Boccaccio<sup>1,3</sup> · Federica Verginelli<sup>1</sup> · Silvia Benvenuti<sup>1</sup> · Miriam Dellino<sup>4</sup> · Paolo Comoglio<sup>5</sup> · Filippo Montemurro<sup>1</sup> · Elena Geuna<sup>1</sup> · Caterina Marchiò<sup>1,2</sup> · Anna Sapino<sup>1,2</sup>

Received: 22 July 2022 / Revised: 10 October 2022 / Accepted: 15 October 2022 / Published online: 8 November 2022

© The Author(s) 2022, corrected publication 2022

## Abstract

The aim of this study is to envisage a streamlined pathological workup to rule out CUPs in patients presenting with MUOs. Sixty-four MUOs were classified using standard histopathology. Clinical data, immunocytochemical markers, and results of molecular analysis were recorded. MUOs were histologically subdivided in clear-cut carcinomas (40 adenocarcinomas, 11 squamous, and 3 neuroendocrine carcinomas) and unclear-carcinoma features (5 undifferentiated and 5 sarcomatoid tumors). Cytohistology of 7/40 adenocarcinomas suggested an early metastatic cancer per se. In 33/40 adenocarcinomas, CK7/CK20 expression pattern, gender, and metastasis sites influenced tissue-specific marker selection. In 23/40 adenocarcinomas, a “putative-immunophenotype” of tissue of origin addressed clinical-diagnostic examinations, identifying 9 early metastatic cancers. Cell lineage markers were used to confirm squamous and neuroendocrine differentiation. Pan-cytokeratins were used to confirm the epithelial nature of poorly differentiated tumors, followed by tissue and cell lineage markers, which identified one melanoma. In total, 47/64 MUOs (73.4%) were confirmed CUP. Molecular analysis, feasible in 37/47 CUPs (78.7%), had no diagnostic impact. Twenty CUP patients, mainly with squamous carcinomas and adenocarcinomas with putative-gynecologic-immunophenotypes, presented with only lymph node metastases and had longer median time to progression and overall survival (< 0.001), compared with patients with other metastatic patterns. We propose a simplified histology-driven workup which could efficiently rule out CUPs and identify early metastatic cancer.

**Keywords** Malignancies of undefined primary origin (MUO) · Cancer of unknown primary (CUP) · Immunohistochemistry · Tissue of origin · Carcinoma

---

Alberto Pisacane and Eliano Cascardi contributed equally to this work.

---

\* Eliano Cascardi  
eliano.cascardi@ircc.it

<sup>1</sup> Candiolo Cancer Institute, FPO-IRCCS,  
10060 Candiolo, Turin, Italy

<sup>2</sup> Department of Medical Sciences, University of Turin,  
10100 Turin, Italy

<sup>3</sup> Department of Oncology, University of Turin Medical  
School, 10060 Candiolo, Turin, Italy

<sup>4</sup> Department of Biomedical Sciences and Human Oncology,  
University of Bari “Aldo Moro”, Bari, Italy

<sup>5</sup> IFOM, FIRC Institute of Molecular Oncology, 20019 Milan,  
Italy

## Introduction

The finding of a metastatic cancer at the time of first diagnosis is a dramatic event for patients and at the same time a frustrating experience for clinicians. It is currently estimated that, despite many efforts, in about 3% of metastatic patients, the tissue of origin of the neoplastic lesion remains unknown; hence, the patient is regarded as harboring a cancer of unknown primary (CUP) [1]. However, CUP diagnosis requires an extensive clinical, instrumental, and pathological workup before being confirmed [2]. The NICE guidelines distinguish indeed “Malignancy of Undefined primary Origin (MUO)” and “provisional” CUP that is “confirmed” only if “no primary site is detected despite selected initial screen of investigations, specialist review, and further specialized investigations as appropriate” [2]. Regrettably, when receiving cases in consultation or patient-initiated second opinion, the pathologists are not



always informed whether the case is a MUO or a provisional CUP. The Royal Collage of Pathologists [3], aligning with NICE recommendations, declared crucial a stepwise approach that uses clinical context, morphology, immunohistochemistry, and, occasionally, other techniques including molecular analysis to confirm or exclude CUP. Gene expression profiling [4] and epigenetic profiling [5] have been introduced de facto with the purpose of unmasking the tissue of origin. Although these studies provided encouraging data, the implementation of such methodologies is far from being prime time in clinical practice, and pathologists are still expected to find the tissue of origin using standard immunohistochemistry (IHC). This pressure leads pathologists to look for the tissue of origin by testing many different tissue-specific markers, which frequently do not resolve the diagnostic problem and waste small specimens, which should otherwise be preserved for high throughput molecular analyses possibly useful for therapeutic decision. The Royal Collage of Pathologists stressed the importance of the histotype to address immunohistochemical analyses correctly. The histotypes considered by ESMO guidelines [6] are well- and moderately-differentiated adenocarcinomas, squamous cell carcinomas, carcinomas with neuroendocrine differentiation, poorly differentiated carcinomas (including poorly differentiated adenocarcinomas), and undifferentiated neoplasms. The same guidelines [6] indicate cytokeratin-7 (CK7) and cytokeratin-20 (CK20) as primary markers in the basic immunohistochemical workup of CUPs, followed by other tissue-specific markers to recognize the tissue of origin. However, recently, Pauli et al. [7] highlighted the challenge in identifying patients with CUP according to ESMO guidelines. Taking all these issues in mind, we decided to consider the real-world approach followed by pathologists to confirm or to exclude a CUP diagnosis and examine how much standard histology can be of aid in this process. Using the classical cyto-histological criteria of different carcinoma histotypes, we reclassified a retrospective series of MUOs and recorded the IHC tests used to study the tissue of origin. We recognized that adenocarcinomas may show some “flavor” of tissue of origin representative of putative immunophenotypes. These latter are suitable to direct targeted investigations to rule out early metastatic cancer. Finally, we analyzed whether CUP histotypes and putative immunophenotypes of adenocarcinomas in the context of distinct clinical presentations could be associated with specific pathways of disease progression and patient survival.

## Materials and methods

### Cohort

Sixty-four patients 38 women (59.4%) and 26 men (40.6%) with an MUO diagnosis were referred to the Candiolo

Cancer Institute FPO-IRCCS (Candiolo, Italy) between 2013 and the first trimester of 2021. The samples were obtained using cytological procedures, core biopsies, and excisional biopsy. One patient (AGN43) gave pre-mortem consent to warm autopsy, and multiple samples were obtained [8]. Patients with hematological malignancies were excluded a priori. Symptom-directed endoscopy and additional imaging were performed if required, according to guidelines [9]. Of note, 34/64 patients performed also PET analysis together with other imaging procedures (CT scans). The site of metastases at disease diagnosis was recorded. PET analysis was never resolute in suggesting the most likely site of origin of the tumor. If only lymph node metastases were present, the number of lymph nodes (single or multiple), the specific lymph node affected, and the lymph node regional localization whether superficial or deep and sub- or supra-diaphragmatics were recorded.

The study was approved by the local ethical committee. Written informed consent was obtained by all patients (AGNOSTOS PROFILING n 010-IRCC-10IIS-15 and upgrade, approved by the ethical committee of FPO-IRCCS Candiolo Cancer Institute). Patients with diagnosis of confirmed CUP were proposed to be enrolled in AGNOSTOS trial (EudraCT Number: 2014–005,018–47), a phase II, randomized, multicenter study to assess the efficacy of nab-paclitaxel-based in association with gemcitabine or carboplatin as first-line therapy.

### Histological and immunohistochemical (IHC) analyses

Slides were reviewed by two pathologists (AP and EC). Similarly to other grading system [9], all cases were graded adopting a scoring system applied to the following parameters: (i) gland formation in any of the architectural variants (e.g., tubular, alveolar, acinar, follicular, papillary) or intracellular mucin in adenocarcinoma, keratinization features in squamous carcinomas, and nests or insular or trabecular growth of cells in NE cancers (score 1: > 75%; score 2: 10–75%; score 3: 1–9%); (ii) mitotic index evaluated in 10 fields of 1 mm<sup>2</sup> (score 1: 0–5 mitoses; score 2: 6–10 mitoses; score 3: > 11 mitoses); and (iii) nuclear pleomorphism defined as variation in size and shape of nuclei (score 1: minimal nuclear pleomorphism; score 2: moderate nuclear pleomorphism; score 3: marked nuclear pleomorphism). The cut-offs for grading were set as follows: grade 1 score 1–5; grade 2 score 6–7; grade 3 score 8–9.

When cases were received for second opinion with IHC for tissue-specific markers already performed, the slides were reviewed by the local dedicated pathologist (AP), and the IHC was repeated, or new markers were tested if deemed necessary, and the residual material was adequate. Nuclear, cytoplasmatic, or membrane expression of tissue-specific

markers and of cell lineage markers were scored, independently from staining intensity, as follows: score 0 = < 5% of neoplastic cells stained; score 1 = 5–10% of neoplastic cells stained; score 2 = 11–50%; and score 3 > 50% of neoplastic cells stained.

## Molecular analysis

When we initiated the AGNOSTOS trial, we offered to CUP patients the ONCOCARTA Gene Panel version 0.1 (Agena Bioscience, Hamburg, Germany) that includes known drug-gable hotspot mutations of *ABL1*, *AKT1*, *AKT2*, *BRAF*, *CDK4*, *EGFR*, *ERBB2*, *FGFR1*, *FGFR3*, *FLT3*, *HRAS*, *JAK2*, *KIT*, *KRAS*, *MET*, *NRAS*, *PDGFA*, *PIK3CA*, and *RET* genes.

Cellularity was the parameter considered for specimen adequacy. DNA was extracted GeneRead DNA FFPE Kit (Qiagen, Hilden, Germany) from four 7- $\mu$ m-thick slides that were hematoxylin and eosin stained and microdissected under a stereomicroscope to enrich for tumor cell content. DNA samples were quantified with spectrophotometric (Nanodrop 1000, Thermo Fisher Scientific, Waltham, MA, USA) and fluorometric (Qubit, Thermo Fisher Scientific, Waltham, MA, USA) assays. Mutational analysis was performed by using the mass spectrometry matrix-assisted laser desorption ionization time-of-flight method with the MassARRAY System (Agena Bioscience, Hamburg, Germany) and the ONCOCARTA as reported above.

## Statistical analysis

Data were analyzed with IBM SPSS Statistics, Version 20.0. (IBM Corp Armonk, NY, USA). Contingency tables with the chi-squared test were applied to infer proportions between groups. Overall survival (OS) was evaluated by

the Kaplan–Meier method and analyzed with the log-rank test. Surviving patients were censored at the date of the last follow-up. A  $p < 0.05$  was considered statistically significant.

## Results

### MUO histotype definition

Tumor histotype, standard morphological parameters, and grade of the 64 MUOs are reported in Table 1. Fifty-four (84.4%) cases showed clear-cut cytohistological features of carcinoma. Of these, 40 (74%) were classified as adenocarcinoma, 11 squamous (20.3%), and 3 neuroendocrine carcinomas (5.5%). Ten cases (15.6%) did not show clear-cut features of epithelial malignancy; 5 of these (50%) were classified as undifferentiated tumors and 5 as sarcomatoid tumors (50%).

Of the 54 cases defined at histology as carcinomas, the majority were either G3 (53.7%) or G2 (37%), while 9.3% were G1 ( $p < 0.01$ ) compared to the undifferentiated and sarcomatoid tumors, which were consistently classified as G3.

### Immunohistochemical workup

#### MUOs with clear-cut carcinoma histology

Adenocarcinomas were the most represented ( $n = 40/64$ , 62.5%) and required relatively extensive immunohistochemical analysis to formulate the diagnosis of provisional CUP and then exclude the presence of a primary tumor by imaging to reach the diagnosis of confirmed CUP [2, 10]. Women and men were 27 (67.5%) and 13 (32.5%), respectively, with a trend of significance for higher women representation compared with the non-adenocarcinoma CUPs ( $p = 0.087$ ).

**Table 1** Histotype, architecture, cell features, and grade of 64 MUO

No of cases (%)	Histotypes	Architectural and cell features	G1 (%)	G2 (%)	G3 (%)
54 (84.4)	Clear-cut features of carcinoma				
40 (74)	Adenocarcinoma	Glandular structures in any of the architectural variants (tubular, alveolar, acinar, follicular, papillary) or intracellular mucin	6 (15)	15 (37.5)	19 (47.5)
11 (20.3)	Squamous cell carcinoma	Solid clusters, sheet-like patterns. Variable keratinization and absence of gland-like structures	0 (0)	4 (36.4)	7 (63.6)
3 (5.5)	Neuroendocrine carcinoma	“Organoid” growth patterns (i.e., trabecular, insular, or sheet-like patterns either singly or in combination). Peripheral palisading may be seen. Round or oval nuclei with “salt and pepper” chromatin	0 (0)	0 (0)	3 (100)
10 (15.6)	No clear-cut features of carcinoma				
5 (50)	Not otherwise specified (NOS)/undifferentiated tumor	Solid growth of epithelioid cells, absence of gland-like structures or keratinization	0 (0)	0 (0)	5 (100)
5 (50)	Sarcomatoid tumors	Solid growth of elongated or pleomorphic cells	0 (0)	0 (0)	5 (100)

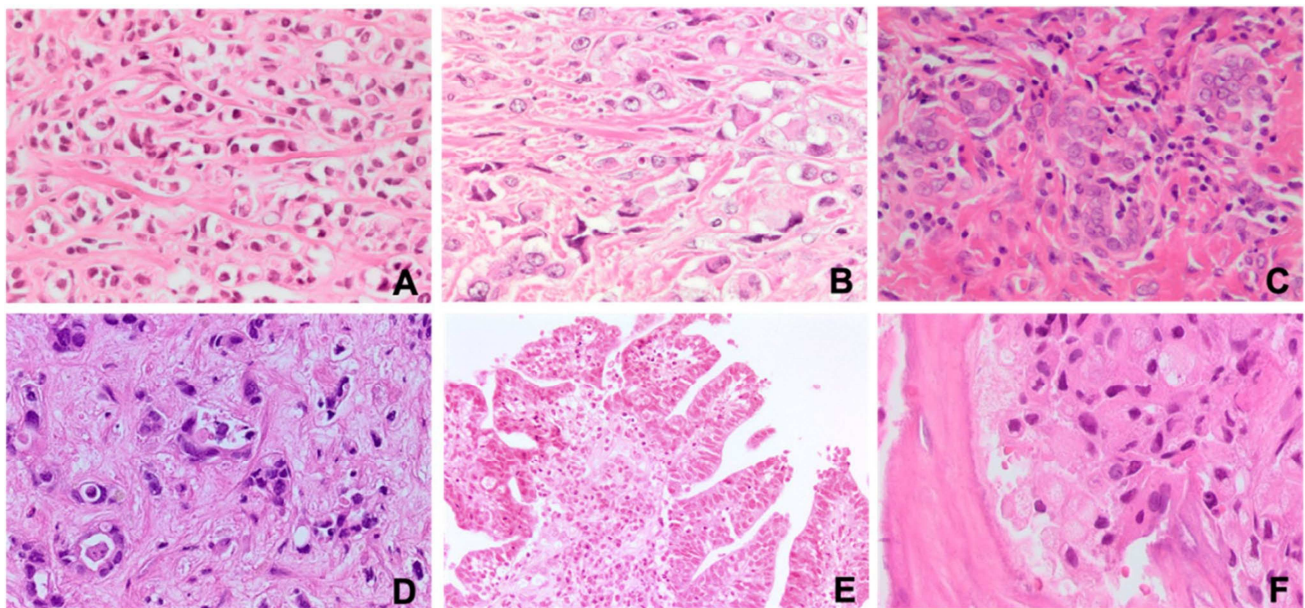


Histological revision identified 7/40 (17, 5%) early metastatic adenocarcinomas. Two lobular breast cancers (AGN40 and EM01/14.15, liver (Fig. 1A) and ovary (Fig. 1B) metastases, both resulted CK7 + /CK20 – and estrogen receptor (ER) score 3. Microfollicular structures were identified on deeper histological sections of a single metastasis in a latero-cervical lymph node, diagnosed at another institution as CUP, TTF1 + (CUP17/8.15) (Fig. 1C). Thyroglobulin expression further confirmed the thyroid origin. AGN336 patient was incidentally discovered with lung and liver localizations after a CT scan for SARS-CoV2 infection. The liver core biopsy was sent to our institution for consultation as “adenocarcinoma ER/TTF1-/GATA3/CDX2 negative, possible CUP.” Based on the morphology of well-differentiated glands in sclerotic stroma, a diagnosis of “intrahepatic cholangiocarcinoma” (Fig. 1D) with lung metastases was given, and the patient was treated accordingly. The tumor was CK7 + /CK20 – /ER – and DOG-1 score 1. AGN329 patient presented with a single localization within the ileo-psoas muscle; her biopsy was sent to our service for consultation as CUP, vimentin score 1 +, and calretinin – /SOX10 – /CD10 –. Based on morphology (Fig. 1E), we suspected a gynecologic origin, which was confirmed by ER, PAX8, and WT1 positivity. The patient had an endometrial biopsy diagnosed as G1 papillary endometrioid carcinoma of the uterus, and the case was reclassified as early metastatic carcinomas. AGN343 patient had a pseudomyxoma peritonei (CK7 – /CK20 score 3 + /CDX2 score 3 /SABT2 score 1, and

ER – /PAX8 – /WT1 –); she underwent right and transversal colectomy during debulking, but neither primary cancers nor polyps were identified. Pancreas and stomach were not affected by any lesion. The patient was treated as carrying a tumor of colorectal origin KRAS mutated (see Fig. 6, ONCOCARTA result). In AGN52, biopsy of a bone metastasis showed signet ring cells at histology (Fig. 1F). WT1 and CDX2 showed both score 1 expression, and TTF1 was negative. The histology-driven endoscopy detected a primary gastric adenocarcinoma.

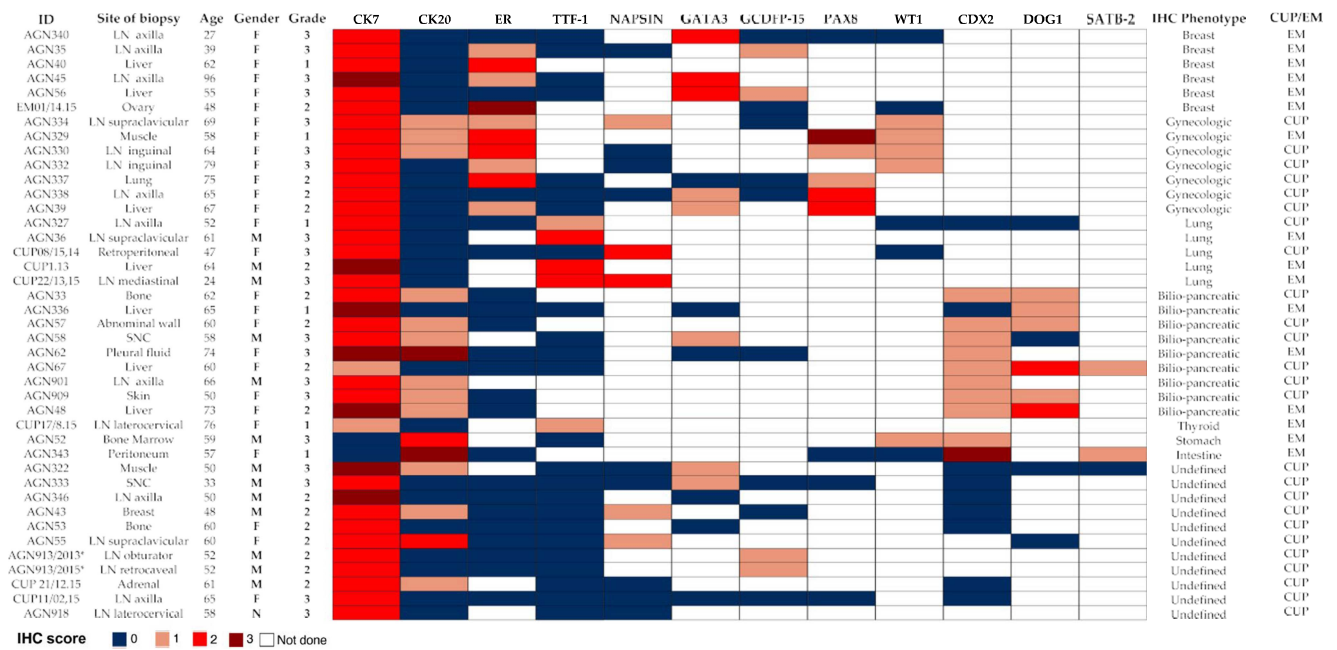
The remaining 33/40 adenocarcinomas were all tested using anti-CK7/CK20 antibodies. Thirty-eight other IHC markers, used in different combinations by pathologists, were recorded in this set. We selected the immunostaining score values of the 10 most recurrent tissue-specific markers to produce a heatmap (Fig. 2) and to evaluate their impact on ruling out CUP diagnosis. A “putative immunophenotype” of potential tissue of origin was suggested based on the results of tissue-specific markers. Except for CK7, the expression score was generally lower in confirmed CUPs than in non-CUP cases.

In 10/33 cases (30, 3%) (3 women and 7 men), none of the markers was convincing for any “flavor” of tissue of origin (undefined immunophenotype), nor any instrumental examination ratified a primary tumor, directly leading to the definition of confirmed CUPs. One of these patients (AGN913) had biopsies of obturator and retrocaval lymph nodes 2 years apart that maintained the same “undefined



**Fig. 1** Cytohistological patterns suggestive of tissue of origin. **A** AGN40 liver and **B** EM01/14.15 ovary biopsies of lobular breast cancer metastasis; **C** CUP17/8.15 latero-cervical lymph node biopsy of follicular thyroid carcinoma; **D** AGN336 liver biopsy of intrahepatic

cholangiocarcinoma; **E** AGN329 ileo-psoas muscle biopsy of G1 papillary adenocarcinoma metastases; **F** AGN52 bone biopsy of signet ring cell gastric cancer metastasis



**Fig. 2** Heat map of tissue-specific marker scores in MUO adenocarcinoma. Heat map depiction of immunohistochemical scores of the most representative tissue-specific markers used in the diagnostic workflow. Data are sorted based on the “putative” immunophenotype (IHC phenotype). Each row represents one case. Columns represent patient identification code (ID), site of biopsy, patient age, patient gender (female: F, male: M), tumor grade (grade 1, grade 2, grade 3),

immunohistochemical scores of tissue-specific markers (IHC score): 0 = < 5% of neoplastic cells stained; 1 = 5–10% of neoplastic cells stained; 2 = 11–50%; 3 = > 50% of neoplastic cells stained; white fields: not performed; confirmed cancer of unknown primary (CUP)/early metastatic cancers (EM). (AGN913)\*two biopsies of the same patients

profile” (CK7 + /CK20 – /ER – /TTF1 – /GCDFP15 score 1). AGN43 has been described previously [8]. During “warm autopsy,” samples of different anatomical sites of metastases were collected, and 16 different tissue-specific markers were performed, but none of them were useful to define the primary tumor.

In the remaining 23/33 (69.7%) adenocarcinomas, all tested for CK7/CK20, a putative immunophenotype was suggested setting the premises for further diagnostic exploration in related organs. Following imaging and/or endoscopy guided examinations, in 9/33 (27.2%) cases, an early metastatic tumor was found, while in the remaining 14/33 (42.4%), the primary was not found, leading to the diagnosis of confirmed CUP.

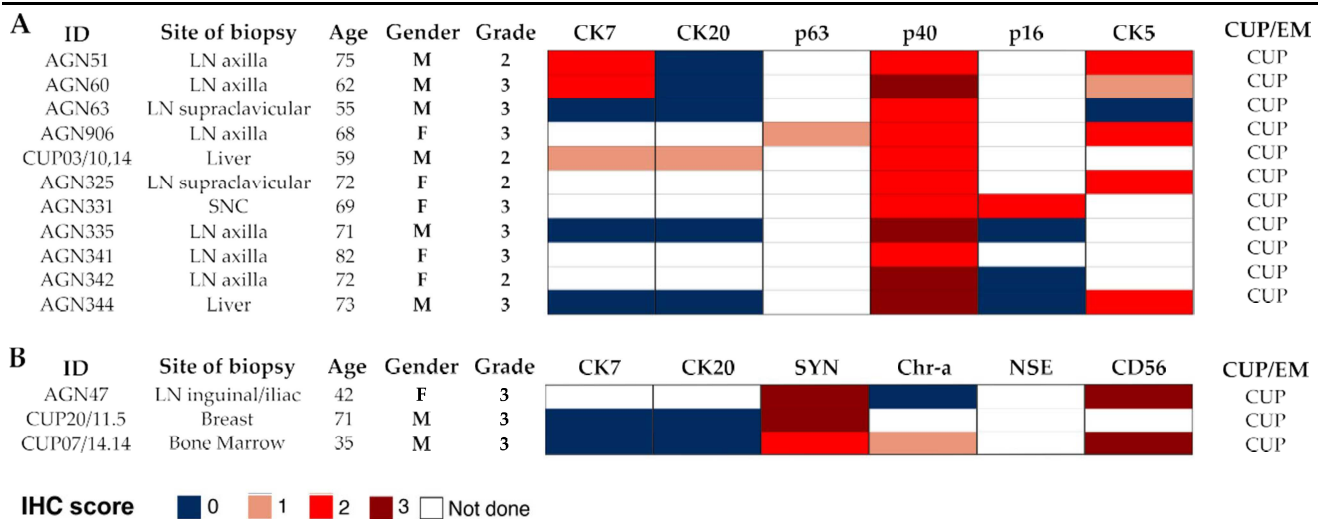
All CK7 + /CK20 – and CK7 + /CK20 + biopsies from women were tested for estrogen receptors (ER). CK7 + /CK20 – /ER + immunophenotype per se, independently from the ER IHC score, identified 2 out of the 6 early metastatic breast cancers. GATA3 was considered as a breast cancer marker in CK7 + /CK20 – /ER – female adenocarcinomas, and other 2 early metastatic breast cancers were identified. In both CK7 + /CK20 – /ER + (1 case) and CK7 + /CK20 + /ER + (2 cases) adenocarcinomas, PAX8/WT1 expression

was performed to rule out gynecological cancers. All cases showing a putative gynecologic immunophenotype were diagnosed as confirmed CUP.

In CK7 + /CK20 – male and in CK7 + /CK20 – /ER – female adenocarcinomas, TTF1 and/or napsin A were the first option, to exclude the lung origin. If TTF1 and/or napsin A were positive (4 cases), pathologists generally did not proceed with other markers. Positron emission tomography (PET) and computed tomography (CT) scan confirmed a possible primary lung cancer in 3 patients, and they were studied for treated consequently.

In CK7 + /CK20 + /ER – and in CK7 + /CK20 – cases, CDX2/DOG1 expression suggested a putative biliopancreatic immunophenotype. Two out of seven cases resulted to be early metastatic cholangiocarcinomas.

In squamous cell and neuroendocrine carcinomas, 24 and 12 different IHC markers were tested, respectively; none were useful to rule out CUP, but only to confirm the cell lineage differentiation of the tumor (Fig. 3A and B). Only the brain biopsy of AGN331 was p16 positive out of 4 squamous carcinomas tested, suggesting a possible HPV-related cancer of the upper respiratory or genital tract, which was not confirmed at further examinations.



**Fig. 3** Heat map of tissue-specific marker scores in squamous and neuroendocrine carcinomas. Heat map depiction of immunohistochemical scores of the most representative tissue-specific markers and cell lineage markers used in **A** squamous carcinoma and **B** neuroendocrine carcinoma. Results are not sorted. Each row represents one case. Columns represent patient identification code (ID), site of

biopsy, patient age, patient gender (female, male), tumor grade (grade 1, grade 2, grade 3), immunohistochemical scores of tissue-specific markers (IHC score): 0 = < 5% of neoplastic cells stained; 1 = 5–10% of neoplastic cells stained; 2 = 11–50%; 3 = > 50% of neoplastic cells stained; white fields: not performed; confirmed cancer of unknown primary (CUP)/early metastatic cancers (EM)

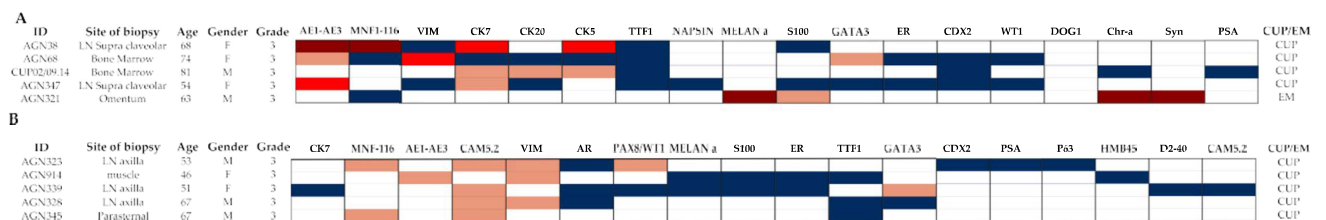
### MUOs with unclear carcinoma histology

Undifferentiated tumors (2 women and 3 men) were investigated with an extensive panel of cytokeratins to confirm their epithelial nature and then with different tissue-specific and cell lineage markers (18 different markers). Excluding the laparoscopic omental biopsy of an undifferentiated tumor expressing Melan A, S100, chromogranin A, and synaptophysin, which led to the endoscopic identification of an intestinal melanoma with aberrant neuroendocrine differentiation, the other cases lacked any tissue-specific marker and were confirmed as CUPs (Fig. 4A). Sarcomatoid tumors were primarily studied with pan-cytokeratins to exclude not epithelial cancers. Fifteen other markers were used; none of them were useful to identify/suggest a tissue of origin (Fig. 4B).

### Histology-driven algorithm for MUO diagnosis

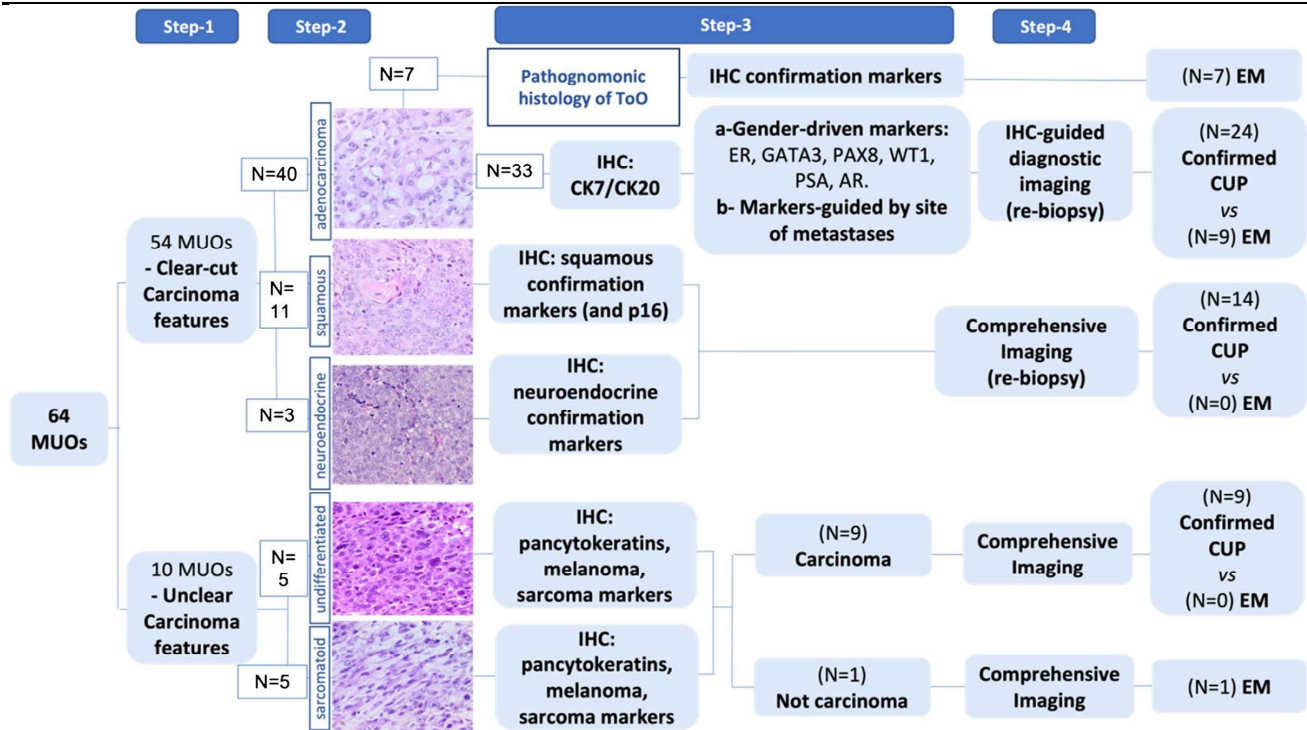
Taking into consideration the results described above, we designed a 4-step workup (Fig. 5).

This algorithm considers “histology” as the guiding parameter of the diagnostic cascade of MUOs, with the intent to define those with clear-cut carcinoma characteristics. This first step followed the definition of the specific tumor histotype in the two sets of MUOs. In adenocarcinomas, cytohistological features per se solved MUOs origin in 7/17 cases, identifying 41% of early metastatic cancers. If no pathological characteristic of a specific organ has been recognized, in adenocarcinomas, the expression of CK7/CK20, along with the sex and site of metastases, headed to the selection of specific IHC markers to define an alleged source tissue immunophenotype, which in turn



**Fig. 4** Heat map of tissue-specific marker scores in undifferentiated and sarcomatoid tumors. Heat map depiction of immunohistochemical scores of the most representative tissue-specific markers and cell lineage markers used in **A** undifferentiated and **B** sarcomatoid tumors. Results are not sorted. Each row represents one case. Columns represent patient identification code (ID), site of biopsy, patient age, patient gender (female, male), tumor grade (grade 1, grade 2, grade 3), immunohistochemical scores of tissue-specific markers (IHC score): 0 = < 5% of neoplastic cells stained; 1 = 5–10% of neoplastic cells stained; 2 = 11–50%; 3 = > 50% of neoplastic cells stained; white fields: not performed; confirmed cancer of unknown primary (CUP)/early metastatic cancers (EM)





**Fig. 5** Workup of MUOs. A 4-step workup is shown. 1st step: differentiation of MUOs in clear-cut and not clear-cut carcinomas. 2nd step: definition of the histotype. 3rd step: selection of specific immunohistochemical (IHC) tissue markers based to the results of step 2. 4th

step: selection of the diagnostic exploration in related organs or comprehensive imaging to confirm the diagnosis of CUP or early metastatic cancer (EM)

guided targeted diagnostic investigations. This pathway led to identify 9/17 (53%) early metastatic adenocarcinomas. In squamous and neuroendocrine carcinomas, the site of metastases and the patient gender were not considered to posit a tissue of origin, and each patient was studied with PET and CT. All were diagnosed as confirmed CUP. The epithelial nature of cancers without clear features of carcinoma was tested using many different pan-cytokeratins and other markers of cell differentiation, which identified one melanoma (5.8% of early metastatic cancers).

### Comparison of clinical and pathological characteristics of confirmed CUP and early metastatic carcinomas

Of the 64 MUOs, 47 (73.4%) were confirmed CUPs, and 17 (26.5%) were early metastatic cancers. Histotype ( $p = 0.045$ ), grade ( $p = 0.04$ ), and putative immunophenotype ( $p = 0.01$ ) were significantly different between confirmed CUP and early metastatic cancers, as shown in Table 2.

### CUP genetic analysis by ONCOCARTA

The oncoprint in Fig. 6 summarizes the genetic analysis by the ONCOCARTA gene panel that explores more than

230 somatic mutations across 19 actionable oncogenes frequently mutated in human cancers.

The leftover of FFPE tissue after the IHC study was adequate for molecular analyses in 37/47 CUP cases (78.7%). Twelve out of the 37 cases (36%) showed at least one mutation; 10 of these were adenocarcinomas, 1 was SSC, and 1 was sarcomatoid carcinoma. The remaining 25 CUPs did not show any detectable alterations. *KRAS* and *PIK3CA* were the most frequently mutated genes and accounted for 75% of all the identified alterations. *KRAS*, *BRAF*, *NRAS*, and *PIK3CA* mutations were found with a mutually exclusive pattern. When subdividing by CUP histological type, the adenocarcinoma was the most frequently mutated (10/12 samples); in addition, only adenocarcinomas harbored *KRAS* mutations. The most mutated immunophenotypes of adenocarcinomas were the undefined (1 *KRAS*; 2 *PIK3CA*, and 1 *BRAF*) and the biliopancreatic (2 *KRAS* and 1 *BRAF*), followed by the gynecological immunophenotype (1 case with concomitant *KRAS* and *PDGFRA* missense mutation and 1 with *PIK3CA* mutation). The pseudomyxoma peritonei with intestinal immunophenotype (AGN343) showed *KRAS* mutation. One sarcomatoid carcinoma (AGN323) showed *NRAS* mutation. One SCC showed co-occurrence of a *PIK3CA* and a *HRAS* mutation. CUPs showing NE differentiation were wild type for the evaluated genes. No specific correlation was found

**Table 2** Clinical and pathological comparison of confirmed carcinoma of unknown primary (CUP) and early metastatic cancers (EM)

		47 CUP		17 EM		<i>p</i> value (95% CI)
		N°	%	N°	%	
Gender	Women	26	55.31	12	70.85	<i>p</i> = 0.52
	Men	21	44.68	5	29.41	
Age	1–34	2	4.25	2	11.76	<i>p</i> = 0.39
	35–70	35	74.46	11	64.70	
	71–96	10	21.27	4	23.52	
Histotype	Adenocarcinomas	24	51.06	16	94.11	<i>p</i> = 0.041
	Squamous carcinomas	11	23.40	0	0	
	Neuroendocrine carcinomas	3	6.38	0	0	
	Sarcomatoid tumors	5	10.63	0	0	
	Undifferentiated tumors	4	8.51	1	5.88	
Grade	Low	1	2.12	5	29.41	<i>p</i> = 0.04
	Intermediate	16	34.04	3	17.64	
	High	30	63.82	9	52.94	
Putative immunophenotype of adenocarcinomas	Biliopancreatic	6	25	3	18.75	<i>p</i> = 0.01
	Breast	0	0	6	37.5	
	Gynecological	6	25	1	6.25	
	Lung	2	8.33	3	18.75	
	Thyroid	0	0	1	6.25	
	Stomach	0	0	1	6.25	
	Intestine	0	0	1	6.25	
	Undefined	10	41.66	0	0	
Site of metastases at diagnosis	Lymph node	21	44.68	3	17.64	<i>p</i> = 0.17
	Lymph node and others*	19	40.42	9	52.94	
	Others*	7	14.89	5	29.41	
Number of metastases	Single	8	17.02	4	23.52	<i>p</i> = 0.45
	Multiple	39	82.97	13	76.47	

\*Others: visceral and/or bone and/or muscle and/or brain site of metastasis

between specific mutations and immunophenotype, sex, histological grade, or survival status.

### Clinical impact of CUP histology and immunophenotype

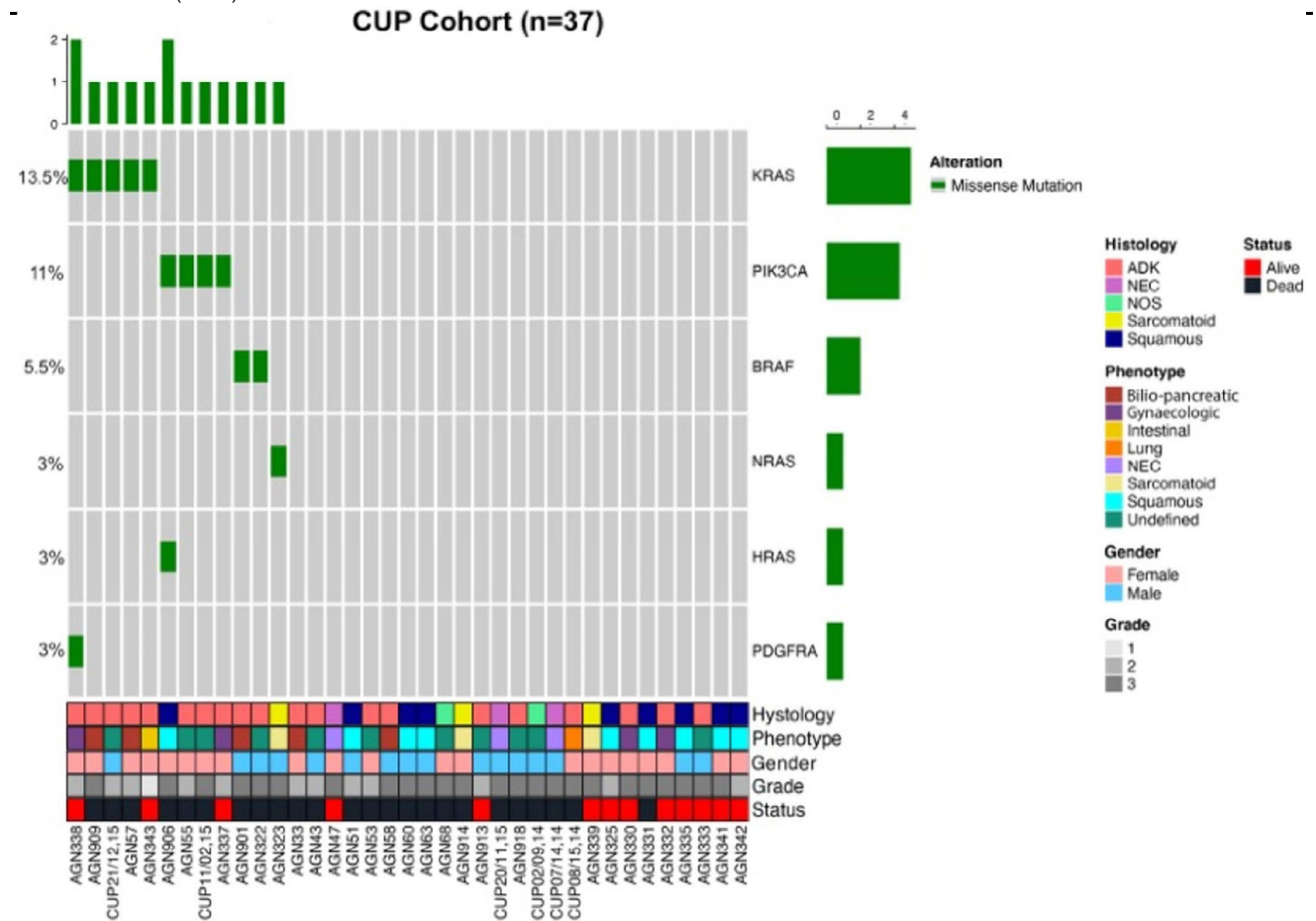
It is known that CUP patients with only lymph node metastases have a better OS [11]. Thus, we focused on confirmed CUP patients and assessed the site of first metastases, the site of progression, and their possible correlation with the histotype and, in the case of adenocarcinomas, with the putative immunophenotype (Tables 3 and 4).

Specifically, squamous cell carcinoma histotype and the “gynecologic immunophenotype” within adenocarcinomas presented frequently with lymph nodes as the solely involved organs. None of the patients within the “bilio-pancreatic” and “lung immunophenotype” of adenocarcinomas presented with lymph node involvement only. Patients with “undefined immunophenotype” of adenocarcinomas, undifferentiated, and sarcomatoid carcinomas frequently progressed, independently from the anatomical site of cancer

at diagnosis. Altogether, 20/47 (42.5%) confirmed CUP patients had only lymph node metastases at first diagnosis, while 27 (57.5%) had metastases elsewhere (visceral and/or bone and/or muscle and/or brain) with or without lymph node involvement (Table 3) (Supplementary Figs. 1A and B). The time to organ progression other than lymph nodes or death for the disease was 6.4 months longer in patients with lymph node metastases only, independently from the number of lymph node involved and the regional location. Finally, the patients with only lymph node metastases at first diagnosis had a significantly longer OS (< 0.001), independently again from the number of lymph node involved and the regional location (sub- or supra-diaphragmatic) (Fig. 7).

### Discussion

Using the data collected from a real-life experience, we here envisage a streamlined pathological workup to rule out CUPs in patients presenting with MUOs.



**Fig. 6** Oncoprint. The graph reports the DNA mutations identified in the cohort of 37 CUPs analyzed by targeted sequencing. Gene names and relative frequency of mutations are reported in the double y-axis. The bar graph in the top of the oncoprint defined the number of vari-

ant/sample. Histology, phenotype, gender, tumor grade, and alive status are annotated at the bottom of the plot. ADK: adenocarcinoma, NEC: neuroendocrine carcinoma, NOS: not otherwise specific or undifferentiated tumors

**Table 3** Evolution of CUP with lymph node localization at diagnosis

Lymph node localization of CUP at diagnosis					No progression/lymph node progression only		Months of FU		Other organs progression/death		Months of FU	
CUP	Total N°	N°	%	N°	Range	Mean	N°	Range	Mean			
<b>Adenocarcinoma immunophenotype</b>												
Gynecologic	6	4	66	4	11.5–78	29.6	0	–	–			
Lung	2	0	0	0	–	–	0	–	–			
Biliopancreatic	6	0	0	0	–	–	0	–	–			
Undefined	10	4	40	1	–	104	3	13–47.2	31.4			
SCC	11	7	70	3	7.4–14.5	10,1	4	18.3–49.7	26.7			
NE carcinoma	3	1	33.3	1	–	65	0	–	–			
Sarcomatoid carcinoma	5	3	60	1	–	9	2	5.3–11	8			
Undifferentiated carcinoma	4	1	25	0	–	–	1	–	14			
<b>Total</b>	<b>47</b>	<b>20</b>	<b>42.5</b>	<b>10</b>		<b>43.5</b>	<b>10</b>		<b>20</b>			

SCC squamous cell carcinoma, NE neuroendocrine

**Table 4** Evolution of CUP with any localization (with or without lymph node) at diagnosis

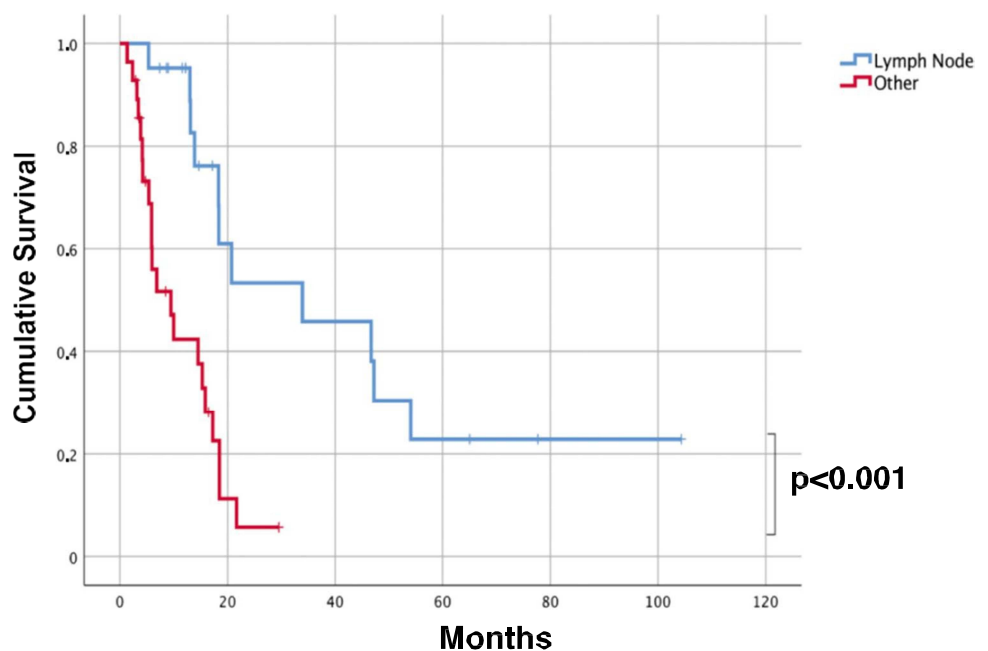
CUP	Total N <sup>o</sup>	N <sup>o</sup>	%	No progression/lymph node progression only N <sup>o</sup>	Months of FU		Other organs progression/death N <sup>o</sup>	Months of FU	
					Range	Mean		Range	Mean
Adenocarcinoma immunophenotype									
Gynecologic	6	2	33	0	–	–	2	1.3–8.5	4.9
Lung	2	2	100	0	–	–	2	6.0–16	11
Biliopancreatic	6	6	100	0	–	–	6	3.2–21	12.8
Undefined	10	6	60	0	–	–	6	0–18.5	6.2
SCC	11	4	30	3	3–19.5	9.3	1	–	50
NE carcinoma	3	2	66.7	0	–	–	2	3.4–18.5	10.9
Sarcomatoid carcinoma	5	2	40	1	–	3.5	1	–	4
Undifferentiated carcinoma	4	3	75	1	–	3.6	2	4.2–14.5	9
Total	47	27	57.4	5		5.4	22		13.6

SCC squamous cell carcinoma, NE neuroendocrine

MUO reviews [12–16] and guidelines [2, 3, 6] are continuously refined, with the aim of defining a diagnostic workup capable of approximating the tissue of origin as closely as possible to better guide therapeutic decisions. With this aim, the CUPISCO trial should have produced a comprehensive genomic profile of patients carrying unfavorable CUPs, to identify possible targets of gene therapy [17]. One of the reasons for the failure of the CUPISCO study was the inadequacy of residual tissue after histopathological diagnosis for high throughput molecular analyses [18]. In the present series, the tissue leftover after immunohistochemical analyses was inadequate for molecular analyses in 21% of the CUP cases.

Another reason of the CUPISCO trial failure was that 10% (13/124) of cases included among unfavorable CUP, at central review turned out as non-carcinoma malignancies [18]. In the present study, undifferentiated and sarcomatoid tumors represented 16.7% of MUOs and were mainly tested with different CK antibody panels to confirm their epithelial differentiation. The sarcomatoid tumors were all positive with at least two pan-cytokeratin antibodies and considered carcinomas, although some positivity for vimentin could be observed. On the other hand, aberrant focal expression of cytokeratin has been reported in melanoma and sarcomas [15]. The single case of melanoma, representing 1.5% of MUOs, was CK-negative and positive

**Fig. 7** Overall survival (OS) analysis. OS survival curve for CUP patients, stratified for the site of the lesion onset (lymph nodes only versus all other sites with or without lymph node involvement). CUP patients who had only lymph node metastases at first diagnosis had a significantly longer overall survival (<0.001) than the patients who had metastases elsewhere (visceral and/or bone and/or muscle and/or brain) with or without lymph node involvement



for Melan-A and S100. In the CUPISCO trial, the misdiagnosis of CUP due to melanomas and sarcomas represented 1.6% and 5.6%, respectively, of the failure cases [18].

The Royal Collage of Pathologists recommends then testing samples with antibodies reactive with germ cell tumors [3]. Accordingly, with NICE guidelines [2], we think that patient age, site of metastases, and serum tumor markers such as alpha-fetoprotein, human chorion gonadotropin, and lactate dehydrogenase [19, 20] may be more effective in the diagnosis of germ cell origin than tissue immunophenotyping.

Once carcinomas are identified, the final step recommended by the guidelines [3] is to determine the likely tissue of origin. Belizzi AM [21] reported that through the application of next-generation immunohistochemistry, pathologists can provide better answers than ever before. The author stressed the role of tumor morphology in leading the CUP diagnostic workflow, together with the acknowledgement that the metastatic site helps tracking the primary lesion. Above all, when lymph nodes are the exclusive site of metastases, it is essential to consider the organs that drain to those lymph nodes as a possible site of origin of the tumor [22]. Actually, the NICE guidelines [2] recommend to manage metastatic squamous carcinoma in cervical lymph nodes as being of head and neck origin and those in inguinal lymph nodes as of anal/lower gynecological tract/urological origin. Our results have shown that patient gender, together with the metastasis site, is an important leading parameter in CUP diagnosis particularly in the set of adenocarcinomas, the most frequent (62.5% of total cases, consistently with previous works [2, 3, 6]) and the most challenging histotype among MUOs. On the other hand, adenocarcinomas may show cytohistological features, for example, follicular and papillary structures or signet ring cells, which per se may be pathognomonic of tumor origin and consequently address tissue targeted markers. In our case series, these features led to identify 42.5% (7/40) of early metastatic adenocarcinomas. If no specific cytohistological patterns are present, in adenocarcinomas CK7/CK20 expression, gender and site of metastases strongly influenced the selection of tissue-specific markers, to depict a “putative immunophenotype” of tissue of origin. All women adenocarcinoma biopsies were in fact tested for ER. In CK7 + /CK20 – cases, ER positivity and GATA3 positivity in ER-negative cases primarily directed clinicians to complete radiological imaging investigations of the breast as a possible site of cancer origin, especially in patients with axillary lymph node metastases. Using these immunophenotypical markers and the cytohistological features crucial for lobular breast cancer diagnosis, 6 early metastatic breast cancers out of 64 MUO cases (9.23%) were confirmed. In the CUPISCO trial, 7.3% of the cases were not compatible with CUP because of proof or strong evidence of a breast primary [18]. In agreement with Pauli C. et al. [18],

the expression of PAX8/WT1 was considered to investigate a possible gynecological origin. If primary ER + breast cancers are excluded, TTF1/napsin A were instead the key tissue markers in male CK7 + /CK20 – and in female CK + /CK20 – ER – MUO biopsies. Based on CUPISCO trial experience [18], although TTF1 expression in a metastatic setting does not unquestionably prove a primary origin in the lung, all TTF1/napsin A positive cases should be referred to radiologists to rule out lung primitivity. In the present series, the “lung immunophenotype” pinpointed 60% of early metastatic lung cancers. Finally, in 25% of the adenocarcinoma cohort, the immunophenotyping did not suggest any tissue of origin (*undefined immunophenotype*) nor any instrumental examination ratified a primary tumor, directly leading to the diagnosis of confirmed CUP.

In squamous cell carcinomas and neuroendocrine carcinomas, the use of cell differentiation markers is advised to avoid misinterpretation, especially when tumor morphology is heterogenous or poorly differentiated, because these histotypes are considered “favorable CUPs” [18].

When we analyzed whether the CUP histotype and the “putative immunophenotype” were associated with specific clinical evolution, we showed that the site of metastases at disease presentation influenced the length of time to progression. Notably, patients with adenocarcinomas of “undefined immunophenotype” and undifferentiated and sarcomatoid carcinomas frequently progressed with diffuse metastases, independently from the anatomical site of cancer at diagnosis, and could be considered in the category of “unfavorable CUPs.” On the other hand, CUP of “the gynecologic immunophenotype” and squamous carcinoma CUP presenting with only lymph node metastases at diagnosis had a longer time to progression. Wach MM et al. [23] showed that nearly two-thirds of patients undergoing axillary or inguinal lymphadenectomy for metastatic squamous carcinoma of unknown primary were alive 5 years following lymphadenectomy. Recently, Pouyiourou M et al. advocate to pursue localized treatment with surgery and/or radiotherapy in single-site and oligometastatic CUP [24]. Although we could not collect the type of treatment of all patients, we showed that CUP patients, who had only lymph node metastases at first diagnosis, had longer time to progression to other organs (6.4 months) and significantly longer overall survival (< 0.001) independently from the number of lymph nodes involved and the regional location (superficial vs deep localization and supra- vs sub-diaphragmatic).

Finally, we recently isolated stem-like cell spheres (agnospheres) from CUP specimens. The agnospheres recapitulated the IHC and molecular phenotypes of CUP and spontaneously and quickly give rise to multiple metastases after subcutaneous engraftment in mice [25]. Interestingly, the CUP samples that did not engraft stemmed from the two



long survivors (AGN47 *neuroendocrine* carcinoma, OS: 65 months, and AGN913 adenocarcinoma with *undefined immunophenotype*, OS 104 months) both presented at diagnosis with metastases limited to lymph nodes [25].

## Conclusions

Standard histology is then essential to drive MUO's diagnostic workup and, when combined with the "putative-immunophenotype" in adenocarcinomas and the metastatic pattern at disease outset, provides prognostic evidence for patients with CUP.

We acknowledge that our study has some limitations. The cohort was relatively small, and a validation cohort was not available; nevertheless, MUOs are rare entities and a careful clinic-pathological examination of these patients is of utmost importance in order not to misinterpret cases of early metastatic disease as CUPs.

These data may pave the way to further validation studies on the diagnostic workup leading to proper identification of CUPs.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00428-022-03435-z>.

**Author contribution** For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, A.S., A. Pisacane, and E.C.; data collection: E.C. and A. Polidori; methodology, I.S., L.C., M.P., F.V., S.B.; software, E.B.; formal analysis, A.S., E.C., E.B. and C.M.; writing—original draft preparation, A.S. and E.C.; writing—review and editing, E.B., C.B., M.D., F.M., E.G.; funding acquisition, C.B., P.C.; A.S. and C.M. All authors have read and agreed to the published version of the manuscript."

**Funding** This work was supported by AIRC-Italian Association for Cancer Research ("Special Program Molecular Clinical Oncology 5 × 1000," N. 9970); to AS, FPRC 5xMille 2015 MIUR (Futuro). This research also received funding by MIUR project "Dipartimenti di Eccellenza 2018–2022" to CM and Ricerca Corrente Ministero della Salute 2022.

**Data availability** All results are reported within the text or in the supplementary material.

## Declarations

**Ethics approval** The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Candiolo Cancer Institute, FPO-IRCCS (Study Protocol N. 010-IRCC-10IIS-15 and following updates; last update: v2.0-16.10.2018).

**Informed consent** Written informed consent was obtained by all patients.

**Conflict of interest** PMC is a shareholder of Metis. CM reports personal consultancy fees from Roche, Bayer, Daiichi Sankyo, and As-

traZeneca, outside the scope of the present work. FM reports personal fees for consultancy or speaker's engagements from Roche, Novartis, Pfizer, AstraZeneca, MSD, Eli Lilly, Daiichi Sankyo, and SeaGen, outside the scope of the present work.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

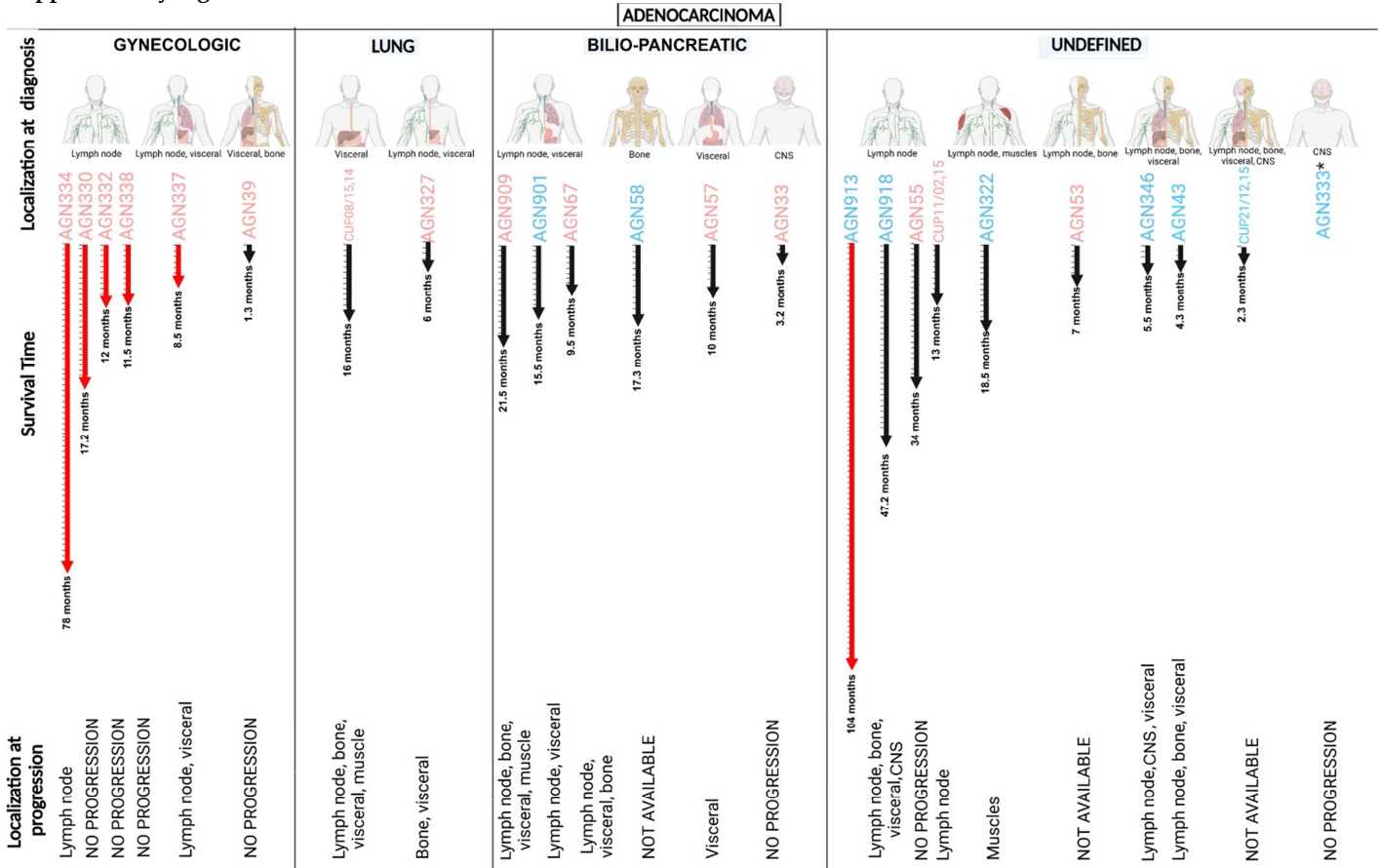
## References

1. Kolling S, Ventre F, Geuna E, Milan M, Pisacane A, Boccaccio C, Sapino A, Montemurro F (2019) "Metastatic Cancer of Unknown Primary" or "Primary Metastatic Cancer"? *Front Oncol* 9:1546. <https://doi.org/10.3389/fonc.2019.01546>
2. National Institute for Health and Care Excellence (2019) Metastatic malignant disease of unknown primary origin in adults: diagnosis and management [NICE Guideline No. 104]. <https://www.nice.org.uk/guidance/cg104>
3. Schofield JB, Oien K (2018) G167 Dataset for histopathological reporting of cancer of unknown primary (CUP) and malignancy of unknown primary origin (MUO)
4. Shen Y, Chu Q, Yin X, He Y, Bai P, Wang Y, Fang W, Timko MP, Fan L, Jiang W (2021) TOD-CUP: a gene expression rank-based majority vote algorithm for tissue origin diagnosis of cancers of unknown primary. *Brief Bioinform* 22:2106–2118. <https://doi.org/10.1093/bib/bbaa031>
5. Moran S, Martinez-Cardus A, Sayols S, Musulen E, Balana C, Estival-Gonzalez A, Moutinho C, Heyn H, Diaz-Lagares A, de Moura MC, Stella GM, Comoglio PM, Ruiz-Miro M, Matias-Guiu X, Pazo-Cid R, Anton A, Lopez-Lopez R, Soler G, Longo F, Guerra I, Fernandez S, Assenov Y, Plass C, Morales R, Carles J, Bowtell D, Mileskin L, Sia D, Tothill R, Tabernero J, Llovet JM, Esteller M (2016) Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol* 17:1386–1395. [https://doi.org/10.1016/S1470-2045\(16\)30297-2](https://doi.org/10.1016/S1470-2045(16)30297-2)
6. Fizazi K, Greco FA, Pavlidis N, Daugaard G, Oien K, Penteroudakis G, Committee EG (2015) Cancers of unknown primary site: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 26(Suppl 5):v133–138. <https://doi.org/10.1093/annonc/mdv305>
7. Pauli C, Bochtler T, Mileskin L, Baciarello G, Losa F, Ross JS, Penteroudakis G, Zarkavelis G, Yalcin S, Ozguroglu M, Beringer A, Scarato J, Mueller-Ohldach M, Thomas M, Moch H, Kramer A (2021) A challenging task: identifying patients with cancer of unknown primary (CUP) according to ESMO guidelines: the CUPISCO trial experience. *Oncologist* 26:e769–e779. <https://doi.org/10.1002/onco.13744>
8. Benvenuti S, Milan M, Geuna E, Pisacane A, Senetta R, Gambardella G, Stella GM, Montemurro F, Sapino A, Boccaccio C, Comoglio PM (2020) Cancer of unknown primary (CUP): genetic evidence for a novel nosological entity? A case report. *EMBO Mol Med* 12:e11756. <https://doi.org/10.15252/emmm.201911756>

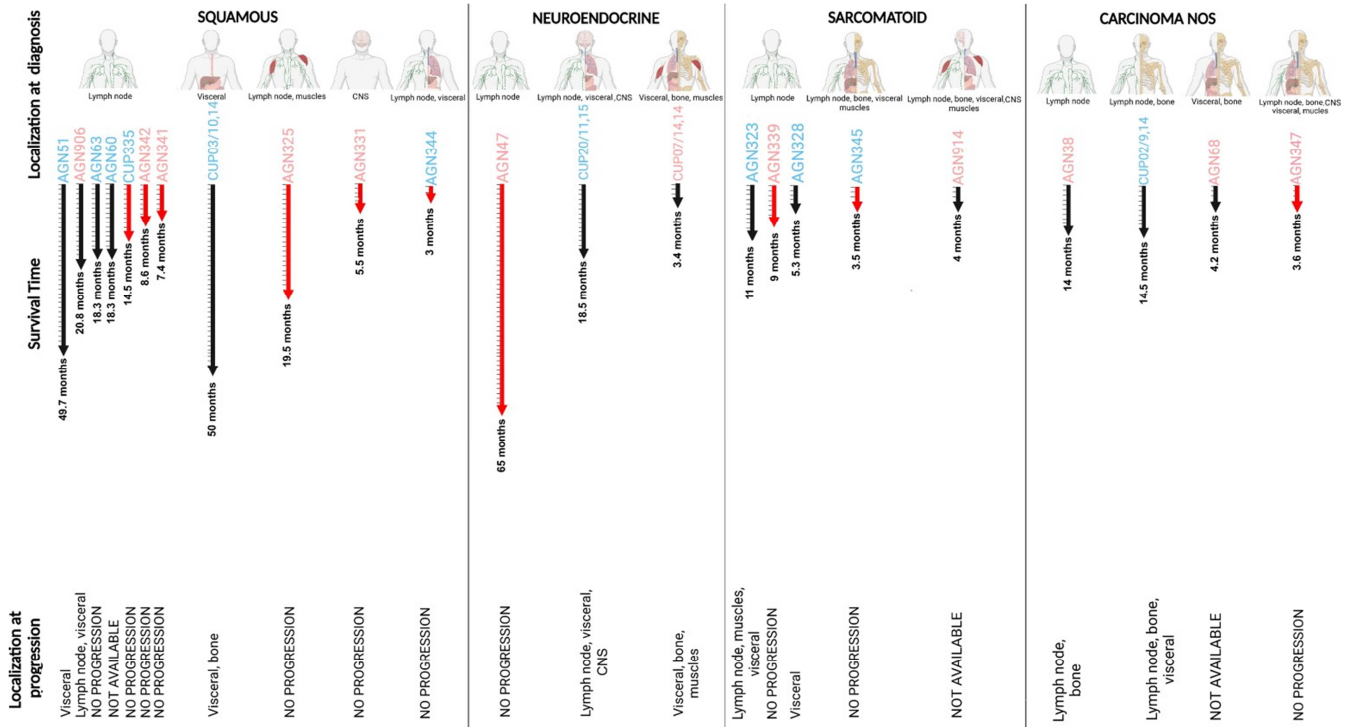
9. Elston CW, Ellis IO (1991) Pathological prognostic factors in breast cancer. I The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 19:403–410. <https://doi.org/10.1111/j.1365-2559.1991.tb00229.x>
10. Cancer NCCf (2010) Diagnosis and management of metastatic malignant disease of unknown primary origin. pp.
11. Hemminki K, Bevier M, Hemminki A, Sundquist J (2012) Survival in cancer of unknown primary site: population-based analysis by site and histology. *Ann Oncol* 23:1854–1863. <https://doi.org/10.1093/annonc/mdr536>
12. Bochtler T, Löffler H, Krämer A (2018) Diagnosis and management of metastatic neoplasms with unknown primary Seminars in diagnostic pathology. Elsevier, 199–206
13. Kato S, Alsafar A, Walavalkar V, Hainsworth J, Kurzrock R (2021) Cancer of unknown primary in the molecular era Trends in cancer
14. Laprovitera N, Riefolo M, Ambrosini E, Klec C, Pichler M, Ferracin M (2021) Cancer of unknown primary: challenges and progress in clinical management. *Cancers* 13:451
15. Lin F, Liu H (2014) Immunohistochemistry in undifferentiated neoplasm/tumor of uncertain origin. *Arch pathol lab med* 138:1583–1610
16. Qaseem A, Usman N, Jayaraj JS, Janapala RN, Kashif T (2019) Cancer of unknown primary: a review on clinical guidelines in the development and targeted management of patients with the unknown primary site *Cureus* 11
17. Ross JS, Sokol ES, Moch H, Mileskin L, Baciarello G, Losa F, Beringer A, Thomas M, Elvin JA, Ngo N (2021) Comprehensive genomic profiling of carcinoma of unknown primary origin: retrospective molecular classification considering the CUPISCO study design. *Oncologist* 26:e394–e402
18. Pauli C, Bochtler T, Mileskin L, Baciarello G, Losa F, Ross JS, Pentheroudakis G, Zarkavelis G, Yalcin S, Özgüroglu M (2021) A challenging task: identifying patients with cancer of unknown primary (CUP) according to ESMO guidelines: the CUPISCO trial experience. *oncologist* 26:e769–e779
19. Gilligan T, Seidenfeld J, Basch E, Einhorn L, Fancher T, Smith D, Stephenson A, Vaughn D, Cosby R, Hayes D (2010) American Society of Clinical O: American Society of Clinical Oncology clinical practice guideline on uses of serum tumor markers in adult males with germ cell tumors. *J Clin Oncol* 28:3388–3404
20. Schmoll H, Souchon R, Krege S, Albers P, Beyer J, Kollmannsberger C (2004) European Germ Cell Cancer Consensus Group. European consensus on diagnosis and treatment of germ cell cancer: a report of the European Germ Cell Cancer Consensus Group (EGCCCG). *Ann Oncol* 15:1377–1399
21. Bellizzi AM (2020) An algorithmic immunohistochemical approach to define tumor type and assign site of origin. *Adv Anat Pathol* 27:114–163. <https://doi.org/10.1097/PAP.000000000000256>
22. Shao Y, Liu X, Hu S, Zhang Y, Li W, Zhou X, Wang Q, Hou Y, Chen Y, Wang Y, Wang Y, Luo Z, Hu X (2020) Sentinel node theory helps tracking of primary lesions of cancers of unknown primary. *BMC Cancer* 20:639. <https://doi.org/10.1186/s12885-020-07042-6>
23. Wach MM, van Beek E, Ayabe R, Ruff S, Brown Z, Goldman DA, Zambirinis CP, Gholami S, Pulitzer M, Hernandez J (2018) Metastatic squamous cell carcinoma of known and unknown primary origin treated with axillary or inguinal lymphadenectomy. *Am J Surg* 216:963–968
24. Pouyiourou M, Wohlfromm T, Kraft B, Hielscher T, Stichel D, von Deimling A, Delorme S, Endris V, Neumann O, Stenzinger A (2021) Local ablative treatment with surgery and/or radiotherapy in single-site and oligometastatic carcinoma of unknown primary. *Eur J Cancer* 157:179–189
25. Verginelli F, Pisacane A, Gambardella G, D'Ambrosio A, Candiello E, Ferrio M, Panero M, Casorzo L, Benvenuti S, Cascardi E, Senetta R, Geuna E, Ballabio A, Montemurro F, Sapino A, Comoglio PM, Boccaccio C (2021) Cancer of unknown primary stem-like cells model multi-organ metastasis and unveil liability to MEK inhibition. *Nat Commun* 12:2498. <https://doi.org/10.1038/s41467-021-22643-w>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Figure 1A



Supplementary Figure 1B









Supplementary Figure 1. Graphical representation of disease progression of patients with (A) adenocarcinomas and (B) CUPs not-adenocarcinoma. Schematic Localization of the tumor at diagnosis for each patient. The timeline scale represents the shortest time interval (40 days) of follow up. The progression and the site of progression is reported. Red timelines: time of follow-up (months) of alive patients; black timelines: time of follow-up (months) of dead patients. \* the patient died after 2 weeks.

## ARTICLE


<https://doi.org/10.1038/s41467-021-22643-w>

OPEN

# Cancer of unknown primary stem-like cells model multi-organ metastasis and unveil liability to MEK inhibition

Federica Verginelli <sup>1</sup>, Alberto Pisacane<sup>2</sup>, Gennaro Gambardella <sup>3,4</sup>, Antonio D'Ambrosio<sup>1</sup>, Ermes Candiello<sup>1</sup>, Marco Ferrio<sup>1</sup>, Mara Panero<sup>2</sup>, Laura Casorzo<sup>2</sup>, Silvia Benvenuti<sup>5</sup>, Eliano Cascardi<sup>2,6</sup>, Rebecca Senetta<sup>2</sup>, Elena Geuna<sup>7</sup>, Andrea Ballabio <sup>3,8,9</sup>, Filippo Montemurro <sup>7</sup>, Anna Sapino <sup>2,6</sup>, Paolo M. Comoglio<sup>5,10</sup> & Carla Boccaccio <sup>1,11</sup>✉

Cancers of unknown primary (CUPs), featuring metastatic dissemination in the absence of a primary tumor, are a biological enigma and a fatal disease. We propose that CUPs are a distinct, yet unrecognized, pathological entity originating from stem-like cells endowed with peculiar and shared properties. These cells can be isolated in vitro (agnospheres) and propagated in vivo by serial transplantation, displaying high tumorigenicity. After subcutaneous engraftment, agnospheres recapitulate the CUP phenotype, by spontaneously and quickly disseminating, and forming widespread established metastases. Regardless of different genetic backgrounds, agnospheres invariably display cell-autonomous proliferation and self-renewal, mostly relying on unrestrained activation of the MAP kinase/MYC axis, which confers sensitivity to MEK inhibitors in vitro and in vivo. Such sensitivity is associated with a transcriptomic signature predicting that more than 70% of CUP patients could be eligible to MEK inhibition. These data shed light on CUP biology and unveil an opportunity for therapeutic intervention.

<sup>1</sup>Laboratory of Cancer Stem Cell Research, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Turin, Italy. <sup>2</sup>Unit of Pathology, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Turin, Italy. <sup>3</sup>Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Naples, Italy. <sup>4</sup>University of Naples Federico II, Department of Chemical Materials and Industrial Engineering, Naples, Italy. <sup>5</sup>Laboratory of Exploratory Research and Molecular Cancer Therapy, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Turin, Italy. <sup>6</sup>Department of Medical Sciences, University of Turin Medical School, Turin, Italy. <sup>7</sup>Multidisciplinary Oncology Outpatient Clinic, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Turin, Italy. <sup>8</sup>University of Naples Federico II, Department of Medical and Translation Science, Naples, Italy. <sup>9</sup>Jan and Dan Duncan Neurological Research Institute, Texas Children Hospital, Houston, TX, USA. <sup>10</sup>IFOM, FIRC Institute of Molecular Oncology, Milan, Italy. <sup>11</sup>Department of Oncology, University of Turin Medical School, Candiolo, Turin, Italy. ✉email: [carla.boccaccio@ircc.it](mailto:carla.boccaccio@ircc.it)



**C**ancer of unknown primary (CUP) is the diagnosis received by patients that present multiple metastases in the absence of a primary tumor anatomically or histologically recognizable through a standardized work-up that includes thorough body imaging and tissue immunohistochemistry<sup>1–4</sup>. Although implying some uncertainties, the definition of CUP applies to 1–2% of all malignancies and features a dismal median survival (<1 year). The clinical course is, however, less aggressive in a subset of cases (15–20%) displaying features reminiscent of some specific tissue of origin, such as the neuroendocrine carcinoma<sup>2,4</sup>.

So far, CUPs were investigated with two main pragmatic clinical aims, such as to uncover the molecular or epigenetic signature of a putative “tissue of origin”, and treat each CUP as a high-grade metastatic tumor of that tissue or organ<sup>5–7</sup>, or to identify mutated cancer genes and inform personalized targeted therapies<sup>8,9</sup>. However, apart from a few exceptions, prediction of a putative origin and application of tailored therapies negligibly extended the overall survival of CUP patients, which remains among the poorest in oncology<sup>3,4,10</sup>. To make progress, it seems critical to adopt an alternative approach, and investigate CUPs as a group of tumors sharing the common ability to disseminate in a way that is (i) early and rapidly progressing; (ii) unrestrained by those barriers in the tissue of origin that facilitate the growth of a detectable primary mass; and (iii) associated with an early cell differentiation block, which precludes formation of recognizable tissues.

In this study, we isolated a panel of human CUP-initiating stem-like cells, named “agnospheres”, able to reproduce a faithful disease model that features early cell dissemination and formation of established multi-organ metastases. We show that agnosphere properties are cell autonomous and converge on constitutive activation of the proliferative MAP kinase pathway, sustaining expression, and activity of the *MYC* proto-oncogene. MEK inhibition with trametinib causes agnosphere cell death, and necrosis of experimental CUP tumors, without inducing negative feedback mechanisms. Importantly, we also show that response to trametinib is foreseen by an originally set up gene expression signature, which, applied to patients’ tissues, predicts sensitivity in a high percentage of CUPs, suggesting that common disease molecular mechanisms are amply shared.

## Results

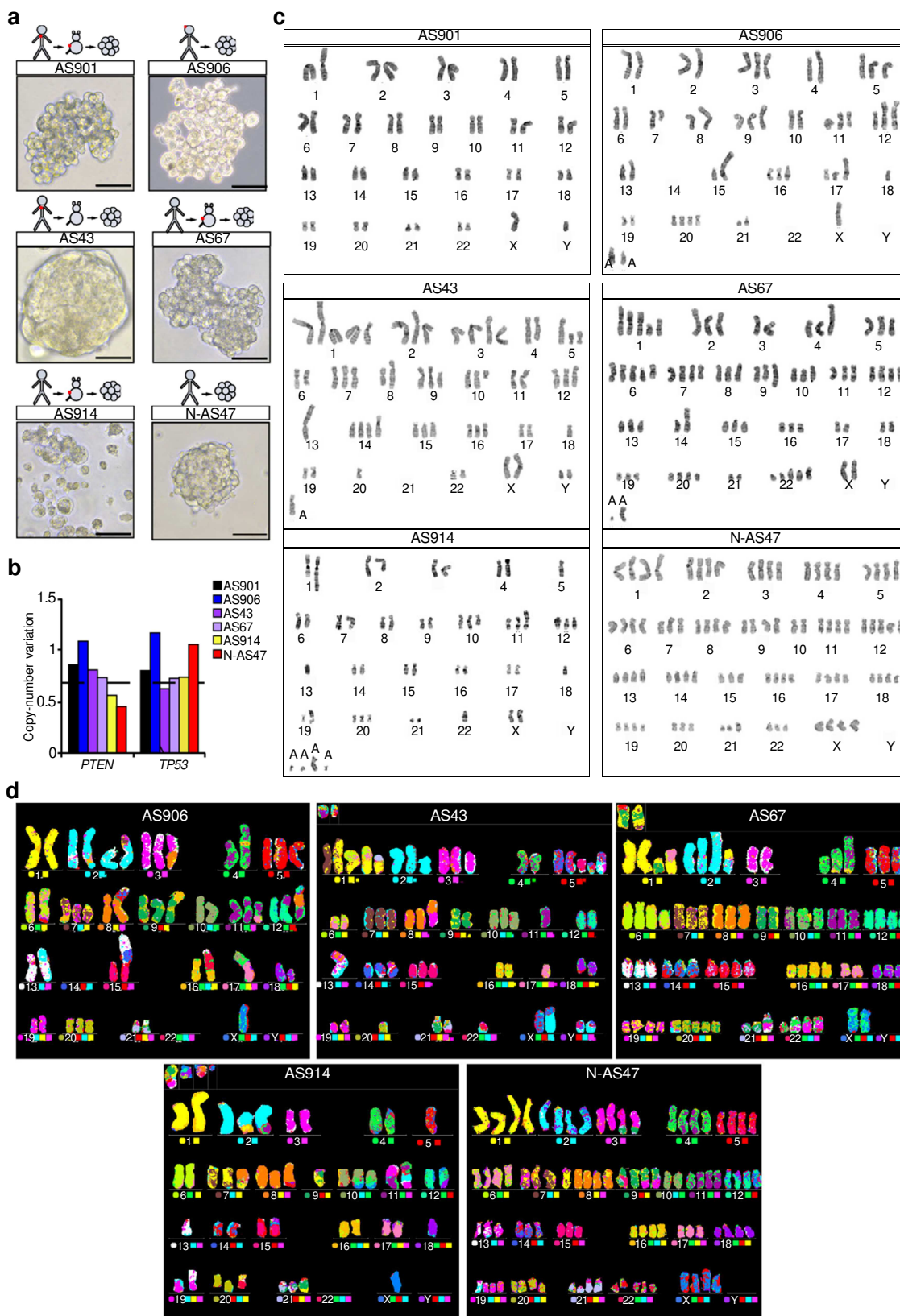
Agnospheres display the genetic makeup of original tumors. In a cohort of 61 early metastatic cancer patients (i.e., that presented at first diagnostic imaging assessment without an evident primary tumor), enrolled at our institution, 27 were diagnosed as CUPs through a rigorous ad excludendum protocol that ruled out the presence of a primary lesion or a defined tissue of origin<sup>2,4</sup>. In the CUP cohort, fresh human specimens for biological studies could be obtained only in eight cases (29%), including five biopsies and three surgeries (Supplementary Tables 1 and 2). All eight samples were transplanted in immunocompromised mice (patient-derived xenografts, PDX), resulting in 6/8 engraftments (75% success rate). Interestingly, CUP samples that did not engraft belonged to a patients’ subset with prolonged survival: AGN47 (OS > 40 m) and AGN913 (OS > 84 m). From PDX, or fresh human samples whenever possible, we attempted spheroid cultures of stem-like cells, named “agnospheres”, with a 75% of successful rate (Fig. 1a and Supplementary Table 2). Briefly, five out of six patients that provided agnospheres (AGN901, AGN906, AGN43, AGN67, and AGN914) displayed the typical aggressive CUP presentation, featuring (i) multiple metastases; (ii) the histological aspect of poorly differentiated carcinomas lacking expression of markers associated with specific organs or tissues, and (iii) rapidly lethal

clinical outcome (average overall survival: 11.5 months; Supplementary Table 1). Agnospheres derived from the above patients are hereafter indicated as AS901, AS906, AS43, AS67, and AS914, respectively. Patient AGN47 presented with multiple metastases but, unlike the above cases, displayed signs of differentiation, leading to the diagnosis of “neuroendocrine CUP” and a better prognosis. The corresponding agnosphere is henceforth indicated as N-AS47.

We performed the genetic characterization of agnospheres and the corresponding available human tissues (hereafter indicated as “original tissues”), by whole-exome sequencing (Supplementary Table 3): AS901 and AS67 resulted hypermutated, consistently with the presence of a *POLE* mutation in AS901, and *POLE* and *POLQ* mutations in AS67; AS906, AS43, and AS914 harbored different combinations of cancer-associated genes. In contrast, N-AS47 harbored no mutation in known oncogenes or tumor suppressor genes (TSG; Supplementary Table 3). Copy-number variation analysis of the two most commonly altered tumor suppressor genes unveiled heterozygous loss of *TP3* accompanied by *TP53* hemizygous mutation in AS43, and heterozygous *PTEN* loss in both AS914 and N-AS47 (Fig. 1b and Supplementary Table 3). The variant allele frequency is consistent with heterozygosity (oncogenes) or hemizygosity (TSG) in agnospheres (Supplementary Table 3). This supports the pathogenic meaning of the genetic alterations and suggests that agnospheres are monoclonal cell populations, at least relatively to driving genetic lesions. Interestingly, we recently showed that multiple synchronous metastases, sampled at warm autopsy of patient AGN43, are highly genetically related and, in particular, share all the driver gene mutations<sup>11</sup>, thus indicating that agnospheres, although derived from a single metastatic lesion, can be representative of all patient’s metastases.

Conventional and spectral karyotypic analysis showed that AS901, consistently with its hypermutated state, did not display evident aberrations of chromosome number and structure (Fig. 1c). In contrast, AS906, AS43, AS914, and AS67 (although hypermutated), exhibited slightly increased ploidy and multiple numerical and/or structural chromosome aberrations, while N-AS47 showed multiple numerical but few structural aberrations (Fig. 1c, d and Supplementary Notes). Although the sample size is too small to generalize conclusions, the overall genetic analysis indicated mutational and karyotypic heterogeneity among agnospheres, consistently with previous observations in 200 CUP patients<sup>8</sup>.

Agnospheres are enriched in self-sustaining stem-like cells. Agnospheres were isolated and propagated in suspension, in highly stringent stem culture conditions<sup>12</sup>. Surprisingly, all agnospheres were capable of self-renewing and long-term propagating at clonal density in the complete absence of any exogenous growth factor, including EGF or FGF, usually required for in vitro isolation and long-term maintenance of the stem phenotype in cancer stem cells from highly aggressive tumors<sup>12–15</sup>. Indeed, in our previous studies, cancer stem cells derived from glioblastomas or colorectal cancer metastases could be in vitro long-term propagated in the absence of exogenous growth factors only in rare and highly aggressive cases, mostly featuring constitutive activation of the growth factor receptor/RAS pathway<sup>15–17</sup>. Among these, for comparison with agnospheres, we choose, as representative carcinoma-derived tumorspheres, colorectospheres mCRC729 and mCRC0155, both generated from colorectal cancer liver metastases, and harboring *IGF-2* gene locus amplification or *KRAS* mutation, respectively<sup>17</sup>. Moreover, we newly generated a growth factor-independent melanosphere (mMS321), from a visceral metastasis with unequivocal melanocytic differentiation in



**Fig. 1** Agnospheres can be isolated from cancers of unknown primary and display the genetic makeup of the original tumors. **a** Morphological appearance of agnospheres derived from CUP tissues propagated in mice or directly from human tissues. **b** *PTEN* and *TP53* copy-number variations measured by qPCR in gDNA from agnospheres. Value = 1 indicates biallelic content; values < 0.7 (dotted line) indicate allelic loss, ( $n = 2$  independent experiments with similar results were obtained). **c** Representative G-banded karyotypes of agnospheres. A: marker chromosome. **d** Representative images of spectral karyotypic analysis (M-FISH) on metaphases of: AS906 ( $n = 16$ ) showing a near-diploid karyotype; AS43 ( $n = 9$ ), containing two major clones with different karyotypes (near-diploid and hypotriploid); AS67 ( $n = 20$ ) with hypertriploid modal karyotype; AS914 ( $n = 20$ ) showing a near-diploid modal karyotype; and N-AS47 ( $n = 13$ ), including three major clones (near-diploid, near-triploid, and near-tetraploid).

the absence of a detectable primary skin lesion, and harboring a *BRAF* mutation (Supplementary Table 1). Independently of exogenous growth factors, all agnospheres displayed the ability to self-sustain the expression of high levels of transcription factors (TFs) known to be major regulators of self-renewal and stem identity in normal and neoplastic tissues, such as Polycomb repressors<sup>18</sup> and reprogramming TFs<sup>19,20</sup>. Specifically, Polycomb repressor EZH2 and BMI1 were highly expressed in the majority of agnosphere cells (Fig. 2a–c and Supplementary Fig. 1a). Among reprogramming TFs, prominent expression of *MYC* gene products, in the absence of gene amplifications (Supplementary Fig. 1b), was observed in all agnospheres (Fig. 2a–c). Co-expression of *MYC* with OCT4, SOX2, and KLF4 was observed only in a fraction of agnosphere cells, likely marking a subpopulation with enhanced stem traits (Fig. 2b and Supplementary Fig. 1a). In tumorspheres, Polycomb and reprogramming TFs were expressed as well, although collectively at a lesser extent, in particular in colorectospheres (Fig. 2c and Supplementary Fig. 1c). Interestingly, YAP/TAZ, recently implied in reprogramming of differentiated cells into stem cells<sup>21</sup>, was highly expressed in all aggressive CUPs, but undetectable in N-AS47 (Fig. 2c).

As the activity of the above TFs is known to drive genetic

programs sustaining the embryonic stem (ES) status, in the overall agnosphere transcriptome we investigated the expression of an “ES cell signature”, based on transcription of 13 gene sets collected in four main groups, including (i) the Polycomb and (ii) the *MYC* target groups, (iii) a group encompassing targets of NANOG, OCT4, and SOX2, and (iv) a group of genes specifically expressed in cultured human ES cells (Fig. 2d and Supplementary Data 1)<sup>22</sup>. By direct comparison, agnospheres and tumorspheres displayed global levels of Polycomb and *MYC* target genes overall similar and comparable to those of ES cells (Fig. 2d and Supplementary Data 1), although N-AS47 and colorectospheres displayed slighter repression of Polycomb targets. Notably, while, in ES cells, maintenance of the transcriptional program sustaining the ES status requires exogenous factors<sup>19,20</sup>, in agnospheres most of the same program is driven in a fully cell-autonomous manner. As the “ES cell signature” was previously shown to be enriched in poorly differentiated aggressive human tumors, and directly correlated with breast cancer tumor grade<sup>22</sup> we investigated its expression in our panel of CUP tissues (including four matched with agnospheres: AGN901, 906, 43, and 47), and, for comparison, in a previously reported panel of breast cancer transcriptomes<sup>22</sup>. CUP tissues displayed an “ES cell signature” globally weaker compared with agnospheres, as expected, but significantly more enriched compared with grade 3 breast cancer tissues (Fig. 2d and Supplementary Data 1), known to contain an abundant fraction of cells with stem properties<sup>13</sup>. Overall transcriptomic data suggest that CUP tissues contain a high proportion of cells that retain stem traits, consistently with their poorly differentiated histology.

Given the interconnection between stem phenotype, metastatic ability, and epithelial–mesenchymal transition (EMT)<sup>23</sup>, we assessed protein expression of EMT core TFs ZEB1 and SNAI2, epithelial markers E-Cadherin and EpCAM, and mesenchymal markers Vimentin and CD44 (Fig. 2c). Moreover, we analyzed the full panel of core and accessory EMT TFs in the agnosphere transcriptome (Supplementary Fig. 1d)<sup>24</sup>. Agnospheres derived from CUP adenocarcinomas (AS901, AS43, and AS67, Supplementary Table 1) and neuroendocrine carcinoma (N-AS47), which maintained in culture an epithelioid phenotype (Fig. 1a), did not display EMT TF or marker expression (Fig. 2c). In contrast, AS906 and AS914, which derived from carcinomas with sarcomatoid features (Supplementary Table 1) and formed loose spheroids in culture (Fig. 1a), expressed ZEB1, SNAI2, CD44, and Vimentin, but negligible levels of EpCAM and E-Cadherin

(CDH1; Fig. 2c and Supplementary Fig. 1e). Accordingly, transcriptomic analysis showed global higher EMT TF expression in AS906 and AS914, while the maximum levels were observed in mMS321 (Supplementary Fig. 1e). This analysis showed also that CUP tissues globally displayed higher EMT TF expression as compared with agnospheres, possibly as result of contamination by tumor microenvironment cells. Validation by tissue immunostaining could not be performed for lack of material, leaving undetermined the real EMT extent in original CUP tissues. Altogether, these findings suggest that EMT seems constitutively activated only in agnospheres derived from CUPs displaying sarcomatoid (i.e., mesenchymal) features. In the other cases, clearly EMT does not occur spontaneously in basic culture conditions, but it could be activated by appropriate exogenous signals and play a key role during metastatic dissemination.

Finally, in the agnosphere and CUP tissue transcriptome, we

also analyzed the expression of cell surface and functional stem cell markers previously used for identification and/or prospective isolation of cancer stem cells from different types of carcinoma. In both agnospheres and CUP tissues, we found significant expression of general markers, such as CD24, CD98, CD166, ITGA6, and ITGB1, with the notable exception of CD133 that was rarely expressed (Supplementary Fig. 1e)<sup>25</sup>. Interestingly, the *MYC* target gene and “don’t eat me” immunosuppressive signal CD47 (refs. 26,27) was broadly expressed (Supplementary Fig. 1e). The overall heterogeneity of cell surface marker expression prevented the design of a common strategy for future prospective isolation of CUP cancer stem cells and further characterization of agnosphere subpopulations.

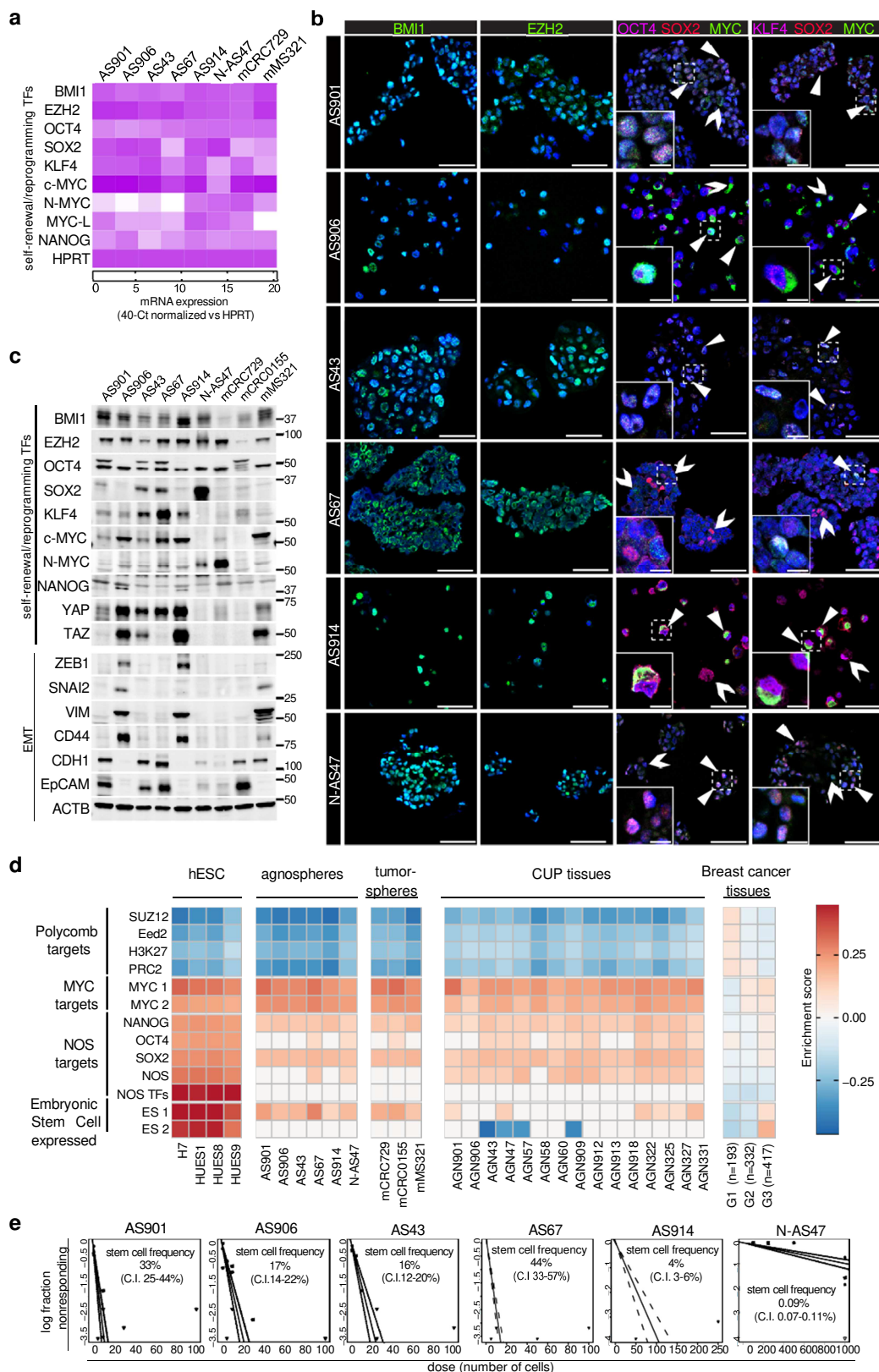
Concerning functional properties, agnospheres kept in the

absence of exogenous growth factors displayed an in vitro estimated fraction of clonogenic cells ranging between 15 and 44% in AS901, AS906, AS43, and AS67, which decreased to 4% in AS914 and to a mere 0.09% in N-AS47 (Fig. 2e). The clonogenic frequency inversely correlated with the population doubling time, which was relatively short in the highly self-renewing agnospheres (2 days in AS901, AS906, and AS67, and 3.5 days in AS43) and increased to ~5 and 10 days in AS914 and N-AS47, respectively.

Overall these data indicate that agnospheres are enriched in cells with transcriptional traits typical of ES cells, mirrored by functional clonogenic properties, which are fully self-sustained in the absence of exogenous growth factors. Such transcriptional features are present also in CUP tissues, attesting to the faithful phenotypic correspondence between the original tumors and the agnospheres, and indicating that CUP tissues are highly enriched in stem-like cells as well.

Agnospheres generate phenocopies of the original tumors and contain a high percentage of tumor-initiating cells. Next, to assess the agnosphere tumorigenic potential, we subcutaneously transplanted 10<sup>5</sup> agnosphere cells in immunocompromised mice (spheropats). At the injection site (IS), tumors formed in all mice, within 20 days in the case of all aggressive agnospheres, and only after 3 months in the case of N-AS47, reflecting the slower progression of the corresponding human tumor (Supplementary Table 1) and the long in vitro cell population doubling time. Importantly, agnospheres regenerated IS-tumors indistinguishable from the original human tumors by histology and expression of immunohistochemical markers (Fig. 3a and Supplementary Fig. 2a), a defining prerogative of cancer stem cells<sup>28</sup>. Interestingly, agnospheres retained in culture the same tissue marker expression, thus behaving as tumoroids (Fig. 3a and Supplementary Fig. 2a)<sup>29</sup>. This property highlights the agnosphere ability not only to proliferate, but also to recapitulate pseudodifferentiative programs





independently of microenvironmental cues. Consistently with the phenotypic and transcriptional correspondence between agno- spheres and original tumors (Fig. 2d), strong and widespread protein expression of Polycomb repressors was displayed by both original and IS-tumors (Fig. 3b and Supplementary Fig. 2b), as well as by agnospheres (Fig. 2b), suggesting that agnosphere

culture conditions did not significantly modify genetic programs and phenotypic traits inherited by original tissues.

In order to evaluate the tumor-initiating potential, we challenged all agnospheres by stringent *in vivo* limiting dilution serial transplantation experiments up to three passages. Remarkably, in all cases but N-AS47, as few as ten agnosphere cells could

Fig. 2 Agnospheres are enriched in stem-like cells that self-sustain their long-term propagation. a Heatmap showing gene expression levels of selected transcription factors (TFs), quantified by qRT-PCR in agnospheres and tumorspheres derived from metastatic colorectal cancer (mCRC729) or melanoma (mMS321). b Representative immunofluorescent stainings of stem cell markers in agnospheres. Arrowheads: marker co-expression; open arrows: cells with single marker expression ( $n = 3$  independent stainings of agnospheres with similar results were obtained). Scale bar, 50  $\mu\text{m}$ . Inset: magnification of dotted area. Scale bar, 10  $\mu\text{m}$ . c Representative western blot analysis of Polycomb repressor factors, reprogramming and EMT TFs, and epithelial and mesenchymal markers in agnospheres and tumorspheres mCRC729, mCRC0155, and mMS321 ( $n = 3$  independent experiments with similar results were obtained, molecular weights are expressed in kDa). d Heatmap showing the gene set enrichment analysis of the 13 gene sets included in the embryonic stem cell signature<sup>22</sup>, performed on the transcriptome of human embryonic stem cell lines (hESC), agnospheres, tumorspheres from metastatic cancers, CUP original tissues, and a panel of breast cancer tissues, pooled by grade (the average enrichment score is shown for each grade). Only significant enrichment scores with a false discovery rate < 10% are shown (Supplementary Data 1). NOS: NANOG, OCT4, and SOX2. e In vitro limiting dilution sphere-forming assay. For each agnosphere, plots generated by the ELDA software are shown, reporting the estimated stem cell frequency (percentage of clonogenic cells) with confidence intervals (C.I.; AS901, AS43, AS67, and AS914:  $n = 3$ ; AS906 and N-AS47:  $n = 4$  independent experiments).

generate tumors after subcutaneous transplantation, determining an estimated tumor-initiating cell (TIC) frequency ranging between 0.3 and 25% at the first passage (P0), and increasing up to 50% in AS901 and AS906 at the third passage (P2; Fig. 3c). Not surprisingly, N-AS47 displayed a relatively low TIC frequency (0.002%), consistently with its overall in vitro biological properties and modest clinical aggressiveness (Fig. 3c). Notably, TIC frequencies in aggressive agnospheres are comparable to the TIC frequency (8.5%) that we measured in melanosphere mMS321 after subcutaneous transplantation (Supplementary Fig. 2c). Melanoma is known as the tumor displaying the highest TIC frequency, also because the subcutis provides an orthotopic environment favoring cell viability and spread<sup>30</sup>. In contrast, tumorspheres derived from other aggressive cancers, when assessed in the same in vivo conditions (subcutaneous transplantation in NOD/SCID mice), display TIC frequencies 1–2 log lower than agnospheres: colospheres up to 0.2% (ref. <sup>31</sup>), breast cancer mammospheres up to 0.1% (ref. <sup>32</sup>), and glioblastoma neurospheres up to 0.1% (ref. <sup>33</sup>).

Altogether these findings highlight the agnosphere ability to recapitulate the original CUP tumors, either by reproducing their histological and pseudodifferentiative features, or by mirroring their clinical aggressiveness in TIC content.

Agnospheres reproduce the multi-organ metastatic pattern of CUP patients. We then assessed metastatic ability, by transplanting subcutaneously luciferase-labeled agnosphere cells from four representative cases (AS901, AS906, AS43, and N-AS47). After removing IS-tumors, to prevent both overgrowth and interference with weaker luminescence from metastatic sites, we longitudinally monitored spheropatiens by in vivo imaging. Increasing luminescent signals were detected in vivo at multiple sites, and further visualized in the explanted organs (Fig. 4a–c). Histopathological and immunohistochemical analysis showed that agnospheres generate metastases phenocopying the original and the IS-tumor (Fig. 4d and Supplementary Fig. 3a–c), indicating that agnospheres retain specific pseudodifferentiative programs rather independent from exogenous signals or tissue contexts, and can adapt to grow in different environments. In spheropatiens transplanted with the aggressive agnospheres (AS901, AS906, and AS43), the metastatic burden lead to the clinical endpoint within 4 months, and the overall postmortem analysis showed multiple metastatic sites in ~75% of mice (Fig. 4e). In N-AS47 spheropatiens, metastatic growth was not fatal, mirroring the patient's clinical course (characterized by a long survival >40 months). However, at the experimental endpoint (7 months), widespread micrometastases and disseminated single cancer cells were detected in 100% of mice, including a remarkable fraction (54% of all mice), where the IS-tumor did not grow (Fig. 4e).

Among the metastatic sites, we found axillary lymph nodes (Fig. 4f), considered as loco-regional lymph nodes invaded by lymphogenous spread, which often occurs in CUP patients as well (Supplementary Table 1)<sup>4</sup>. The other organs colonized in spheropatiens were likely reached by hematogenous spreading, with homing occurring not only in the primary capillary district downstream venous dissemination (lung), but, often, downstream systemic arterial circulation as well. The frequent colonization of connective tissues, including peripancreatic fibroadipose tissue, gonadal adipose tissue, kidney capsule, and the occasional homing between myocardiocytes, suggested a predilection for the connective/mesodermal soil, where micrometastases and intravascular cancerous emboli could be often detected (Fig. 4d and Supplementary Fig. 3a–c). This, again, is consistent with the CUP metastatic pattern, which can include uncommon sites, such as subcutaneous connective tissues and muscles, as observed in several patients, including AGN906, AGN43, AGN914, AGN909, AGN322, and AGN325 in our cohort (Supplementary Table 1). Strikingly, in CUP spheropatiens, metastatic cells efficiently disseminated as early as within 10 days after injection, as observed in organs explanted from 4/4 mice transplanted with AS43. Dissemination occurred well before the IS-tumor became palpable, suggesting that cells can initiate the process without prior local expansion and, likely, without induction of a pre-metastatic niche from an established tumor. Immunostaining of explanted organs showed single disseminated tumor cells, micrometastases and vascular emboli at multiple sites, mostly in the lung and connective tissues around the organs, i.e., the same sites where macrometastases develop (Fig. 4g). Interestingly, single disseminated cells were identified by pan-cytokeratin immunostaining, suggesting that, during metastatic spread, these cells retain their epithelioid phenotype, or reside in a partial EMT state.

Such a rapid and widespread, both lymphogenous and hematogenous, metastatization is rather uncommon when TICs or organoids from other aggressive carcinomas are transplanted subcutaneously (i.e., ectopically), in contrast to neural crest-derived melanoma cells, which, injected orthotopically under the skin, can metastasize<sup>34</sup>. Consistently, we could not observe metastatic dissemination by subcutaneously transplanting mice with colosphere mCRC729, which, as mentioned above, derives from a colorectal cancer metastasis and shares some properties with agnospheres (Fig. 2a–d), in particular cell-autonomous propagation. Indeed, after mCRC729 injection, IS-tumors formed in 100% of cases, being in some cases locally invasive, but always unable to form detectable metastases during 7-month in vivo longitudinal monitoring and ex vivo organ imaging (Fig. 4e and Supplementary Fig. 3d, e).

Overall, these findings attest that agnospheres, in particular those from aggressive CUPs, retain a high spontaneous and multi-organ metastatic ability, which recapitulates the CUP



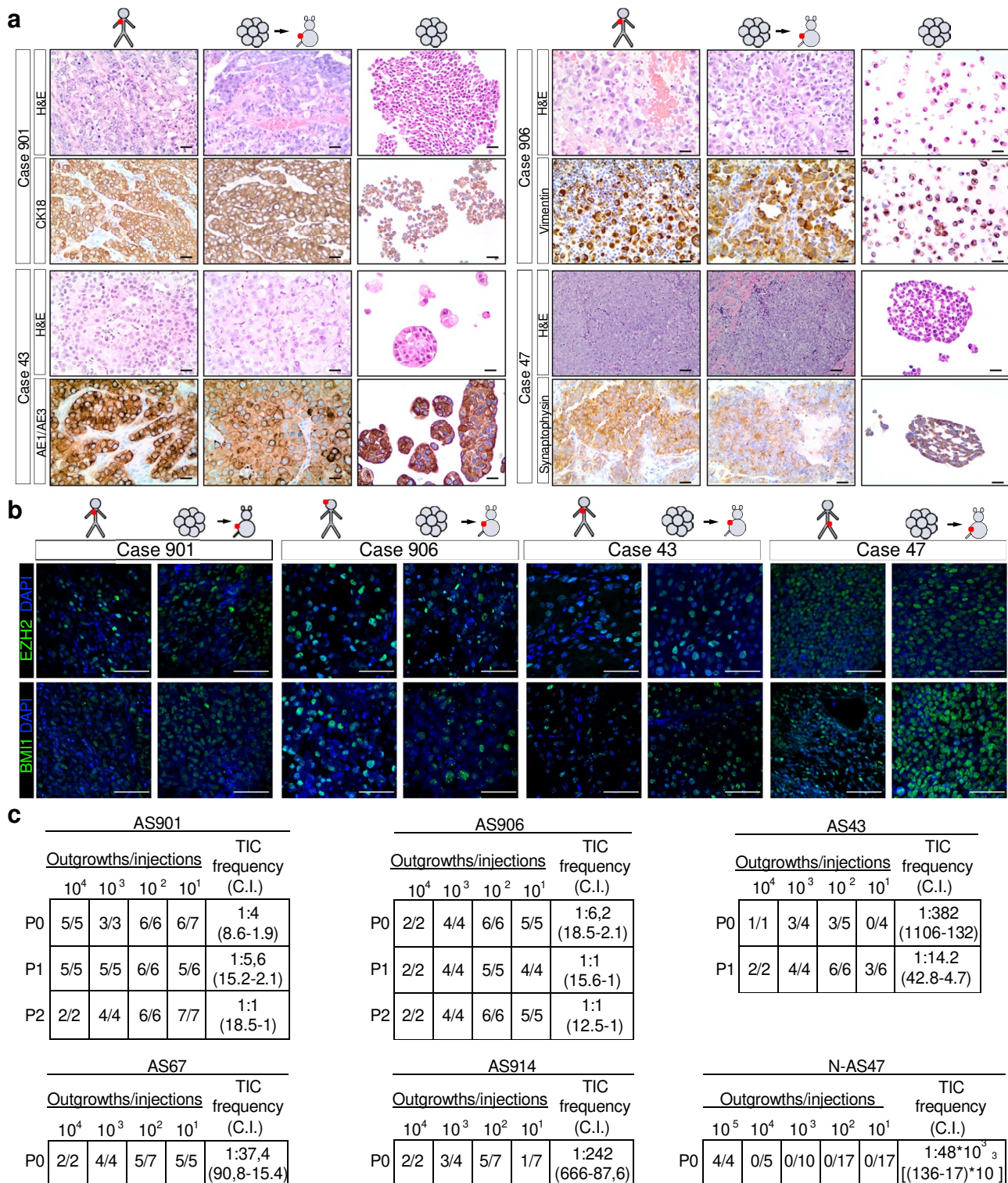
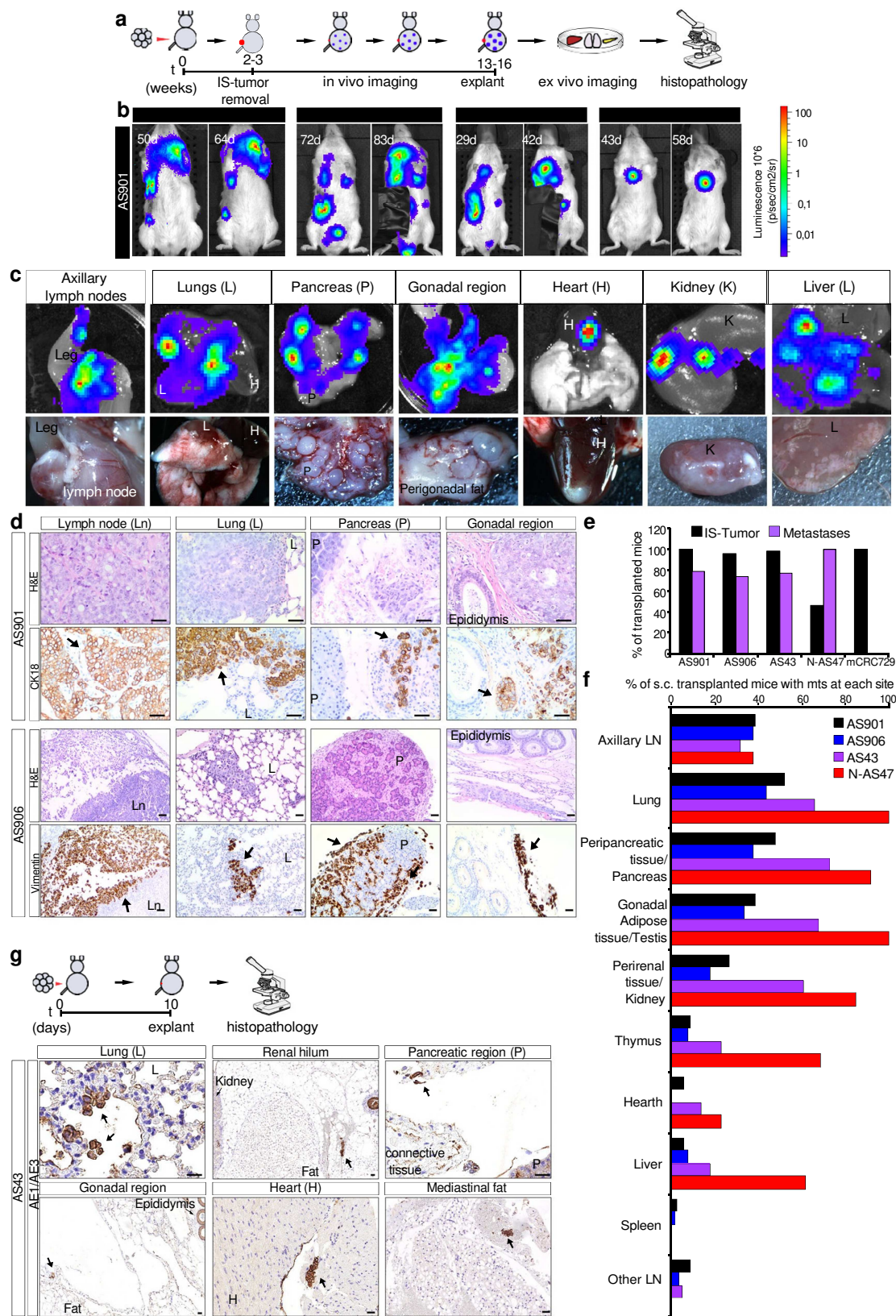


Fig. 3 Agnospheres generate phenocopies of the original tumors and contain a high percentage of tumor-initiating cells. **a** Histopathological features (hematoxylin and eosin staining, H&E) and immunohistochemistry with antibodies used for CUP diagnosis, in patient metastases (left columns), tumors formed by agnospheres at the subcutaneous injection site (middle columns), and cultured agnospheres (right columns). Scale bar, 50  $\mu$ m. A representative experiment is shown ( $n = 3$  independent stainings of agnospheres with similar results were obtained). **b** Immunofluorescent stainings of stem cell markers in patient metastases (left columns) and tumors formed by the corresponding agnospheres at the subcutaneous injection site (right columns). Nuclei were counterstained with DAPI. Scale bar, 50  $\mu$ m. A representative experiment is shown ( $n = 3$  independent stainings of agnospheres with similar results were obtained). **c** In vivo limiting dilution assay. The indicated numbers of agnosphere cells ( $10^5$ – $10^1$ ) were transplanted subcutis into immunocompromised mice (P0) and, where indicated, agnospheres were re-derived in culture and transplanted for a second (P1) or a third (P2) passage. TIC tumor-initiating cell frequency, C.I. confidence interval.





clinical presentation, and may represent a convenient tool to explore every stage of the metastatic cascade.

Agnospheres are sensitive to MEK inhibition in vitro. As described above, all agnospheres shared the distinctive ability to

proliferate and self-renew in the absence of any exogenous growth factor. Moreover, agnospheres, including the slowly proliferating N-AS47, were insensitive to an ample panel of exogenously supplied growth factors, with the exception of AS67, which showed a moderate proliferative response to EGF, NRG1, and FGF (Supplementary Fig. 4a). Agnosphere proliferative autonomy

**Fig. 4** Subcutaneously transplanted agnospheres reproduce the early and multi-organ metastatic pattern of CUP patients. **a** Outline of longitudinal in vivo monitoring of spheropatiens transplanted subcutis with luciferase-labeled agnospheres, followed by endpoint ex vivo organ imaging and histopathological analysis. IS-tumor injection site tumor. **b** Representative in vivo images of mice transplanted with AS901, showing increasing bioluminescent signals at metastatic sites at the indicated time points (d, days). Black tape was applied to mask the IS-tumor residual signal. **c** Representative macroscopic images of the most frequent metastatic sites. Top: bioluminescent signals of different explanted organs; bottom: photographs of macrometastases. **d** Histopathology (H&E) and immunohistochemistry for human CK18 or human Vimentin, revealing metastases in the indicated organs of mice transplanted with AS901 or AS906, respectively. Scale bar, 50  $\mu$ m. Representative experiments are shown (at least  $n = 3$  stainings in independent organs were obtained with similar results). **e** Bar graph showing the percentage of mice transplanted subcutis with agnospheres or colosphere mCRC729 that developed IS-tumors and/or metastases (AS901:  $n = 33$ , AS906:  $n = 50$ ; AS43:  $n = 44$ ; N-AS47:  $n = 13$ ; mCRC729:  $n = 14$ ). **f** Bar graph showing the percentage of mice subcutaneously (s.c.) transplanted with agnospheres as in **e** that displayed metastasis (mts) at each of the listed sites. **g** Immunostainings for pan-cytokeratin AE1/AE3 showing tumor cells disseminated in organs explanted 10 days after AS43 subcutaneous injection. Scale bar, 50  $\mu$ m ( $n = 4$  independent experiments with similar results were obtained).

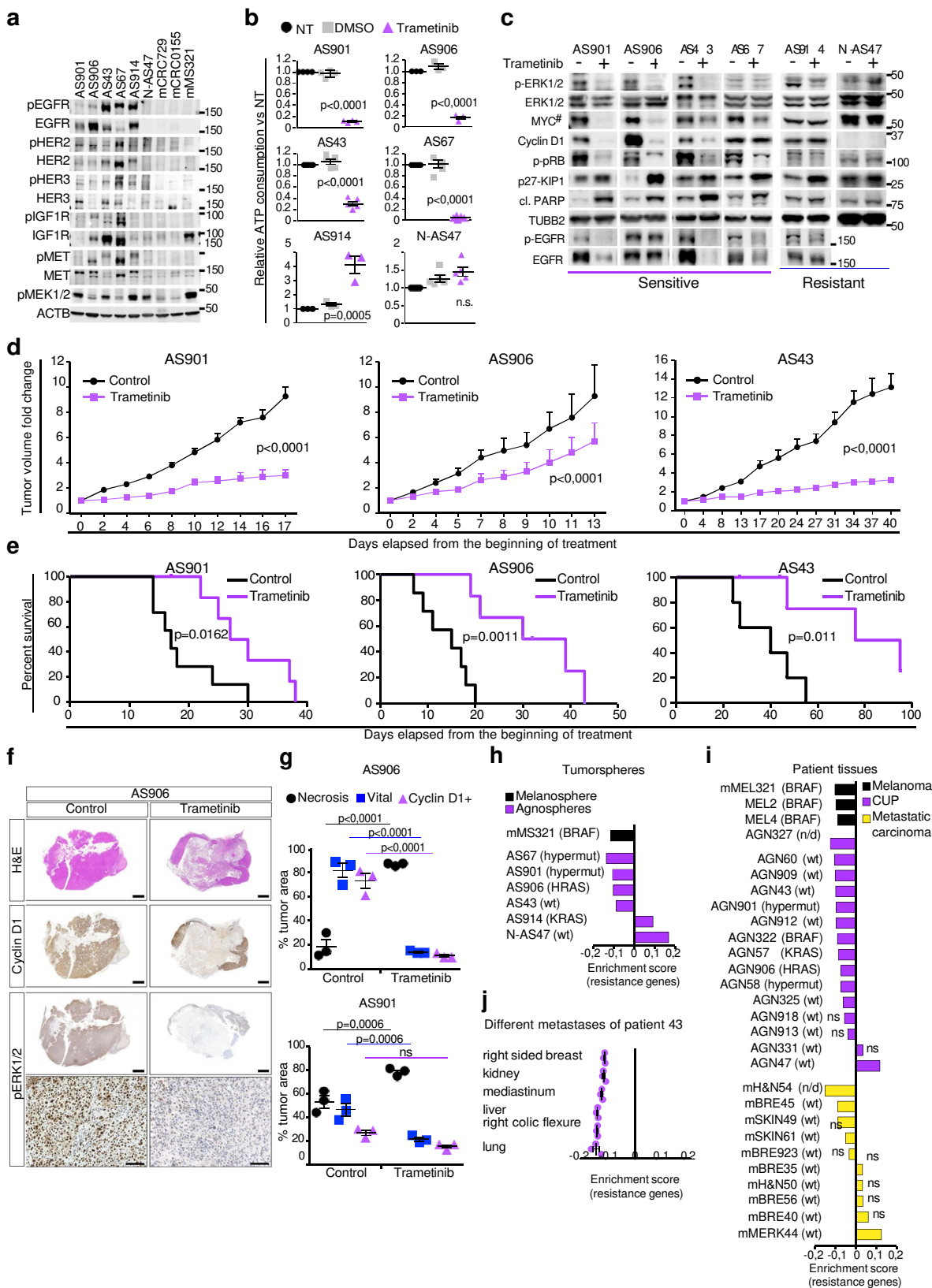
could be associated with the presence of *RAS* oncogenic mutations, known to confer growth independence to tumorspheres<sup>15</sup>, only in the case of AS906 (*HRAS*-mutated) and AS914 (*KRAS*-mutated; Supplementary Table 3). Remarkably, in all agnospheres, proliferative autonomy could be sustained by the expression of several growth factors, among which EGF-like ligands, neuregulins, HGF, and IGF-1/2 (Supplementary Fig. 4b), accompanied by expression and phosphorylation of their genetically unaltered cognate receptors (Fig. 5a), indicating the occurrence of autocrine loops. Conversely, expression and phosphorylation of the same receptors was almost undetectable in tumorspheres that relied on cell-autonomous proliferative triggers, such as BRAF (mMS321) or KRAS (mCRC0155) mutation, or amplification of the *IGF-2* genetic locus (mCRC729)<sup>17</sup>. These findings prompted us to try to halt agnosphere proliferation by blocking constitutively activated receptors, in particular those of the EGFR family, which seemed prominently activated in all aggressive agnospheres. Pan-EGFR family inhibition through lapatinib or afatinib<sup>35</sup> was assessed in three representative agnospheres: AS901 (hypermutated), AS906 (*HRAS* mutant), and AS43 (*RAS* pathway wild type; Supplementary Table 3). A significant proliferative arrest, without cell death induction even after a 14-day treatment, was observed only in AS43 (Supplementary Fig. 4c, d). Indeed, although abolishing EGFR phosphorylation in all agnospheres, EGFR inhibitors partially affected downstream signaling only in AS43 (Supplementary Fig. 4e). Agnosphere partial or full resistance to EGFR inhibition could be explained by the concomitant presence of the *HRAS* mutation (in AS906), as described<sup>15</sup>, and/or other autocrine loops providing bypass mechanisms that kept MAPK and AKT pathways constitutively active (Fig. 5a and Supplementary Fig. 4e).

We therefore reasoned that the multiple mechanisms sustaining agnosphere autonomous proliferation should converge onto the MAPK pathway. As, indeed, agnospheres display high levels of constitutively phosphorylated MEK1/2, in some cases comparable to those of BRAF-mutated mMS321 (Fig. 5a), we challenged agnospheres with two selective, clinically approved, MEK1/2 inhibitors, endowed with different mechanisms of action: (i) trametinib, known to be the most potent MEK1/2 inhibitor, used in BRAF-mutated melanomas<sup>36,37</sup>, or (ii) selumetinib, an earlier inhibitor, known to be less effective than trametinib in fully preventing MEK1/2 activation by BRAF. As assessed by dose–response experiments (Supplementary Fig. 4f), in three representative agnospheres trametinib strongly affected cell viability at nanomolar concentrations, while selumetinib was less effective even at maximal doses (Supplementary Fig. 4f), as reflected by early signaling events, such as ERK1/2 and pRB phosphorylation (Supplementary Fig. 4g). The biological effects exerted in agnospheres by the two inhibitors were comparable to those observed in BRAF-mutated mMS321 (Supplementary Fig. 4h). Trametinib was therefore chosen for further experiments, showing to be effective in 4/6 agnospheres, by inducing dramatic growth arrest within 4 days in AS901 and AS906, and within 10 days in

AS43 and AS67 (Fig. 5b), and concomitant apoptosis, as indicated by the appearance of the cell death effector cleaved-PARP protein (Fig. 5c). In sensitive agnospheres (AS901, AS906, AS43, and AS67), trametinib induced long-term inhibition of ERK1/2 phosphorylation, the direct MEK1/2 target, and downregulation of c-MYC, known to be tightly regulated by proliferative signals via MAPK pathway (Fig. 5c)<sup>38,39</sup>. c-MYC decrease (or disappearance, in the most sensitive agnospheres) was accompanied by proportional downregulation of its transcriptional target Cyclin D1, responsible for G1–S cell cycle progression via CDK4/6-mediated pRB hyperphosphorylation, which was consistently reduced as well (Fig. 5c). Stabilization of the cell cycle inhibitor p27-KIP1, known to be prevented by c-MYC, was also observed (Fig. 5c)<sup>39</sup>. Interestingly, in sensitive agnospheres, trametinib did not cause a rebound increased EGFR phosphorylation and expression, often observed after MEK inhibition and resulting from relief of a negative feedback from MAPK to receptors (Fig. 5c)<sup>37</sup>. Rather, in sensitive agnospheres, EGFR phosphorylation and expression were decreased (Fig. 5d). This suggests that agnospheres may lack this negative feedback from MAPK to receptors, resulting in an unrestrained receptor–MAPK signaling axis, which could become addictive and would eventually cause hypersensitivity to MEK1/2 inhibitors (see “Discussion”). Conversely, trametinib was biologically ineffective in N-AS47 (Fig. 5b), consistently with its inability to prevent ERK1/2 phosphorylation and downstream signaling (Fig. 5c), thereby suggesting that N-AS47 can self-sustain proliferation through pathways independent of MEK1/2, as recently shown in organoids from neuroendocrine tumors<sup>40</sup>. Surprisingly, in KRAS-mutated AS914, trametinib inhibited ERK1/2 phosphorylation but exerted a late paradoxical effect, by increasing cell proliferation (Fig. 5b, c), as described in several cell lines and tumors<sup>37</sup>. This effect was likely a specific outcome of MEK1/2 inhibition, as it was observed after selumetinib treatment as well (Supplementary Fig. 4i). The mechanism of such paradoxical response does not involve EGFR rebound upregulation; however, it likely contributes to keep cMYC and Cyclin D1 highly expressed independently of ERK1/2 inhibition (Fig. 5c).

Trametinib is effective in CUP preclinical models and its efficacy is predicted in patients by a response signature. To evaluate the therapeutic efficacy of trametinib in preclinical models, CUP tumors were generated by subcutaneous transplantation of three representative agnospheres (AS901, AS906, and AS43). After IS-tumor establishment, spheropatiens were randomized and treated with a dose of trametinib (1 mg/Kg/die) previously shown to be well tolerated and effective in xenograft models harboring BRAF mutation (Fig. 5c)<sup>37,41</sup>. We chose to primarily monitor the therapeutic effect in the established IS (subcutaneous) tumor, assuming that (i) the IS-tumor is not representative of a conventional primary tumor, but rather of every metastasis generated by the agnosphere in the mouse, given the homogeneity of IS-tumors and metastases; (ii) the effect of





trametinib is not expected on the early dissemination step of the metastatic process, but on the growth of established metastases; (iii) the IS-tumor can be more accurately and precociously measured than metastases, which are multiple and delayed, and evolve by an unpredictable pattern that prevents reliable longitudinal quantification. The inhibitory effect of trametinib on IS-

tumor growth was statistically significant throughout the experiment (Fig. 5d). Accordingly, the time to reach the experimental endpoint (tumor volume = 1600 mm<sup>3</sup>) was significantly prolonged in all mice treated with trametinib as compared with controls (Fig. 5e). Interestingly, histopathological analysis of tumors collected at the endpoint showed that treated tumors, as

Fig. 5 MEK inhibition is effective in vitro and in CUP preclinical models, and its efficacy is reliably predicted in patients by a gene expression signature. a Representative western blot analysis of phosphorylated (p) and total proteins in agnospheres and tumorspheres ( $n = 3$  independent experiments). b Cell viability of agnospheres treated for 4 days (AS901 and AS906) or 10 days with 50 nM trametinib or vehicle (DMSO) normalized vs. untreated cells (NT; AS901, AS43, and AS67  $n = 4$ ; AS906 and AS914  $n = 3$ ; and N-AS47  $n = 5$  independent experiments, mean  $\pm$  SEM, ANOVA, Bonferroni multicomparison test). c Representative western blot analysis of agnospheres treated 48 h (AS901, AS906, and AS67) or 96 h (AS43, AS914, and N-AS47) with vehicle (-) or 50 nM trametinib (+) (p- phosphorylated, cl. PARP: cleaved PARP, MYC# c-MYC in AS901, AS906, AS43, AS67, and AS914, and N-MYC in N-AS47;  $n = 3$  independent experiments). d, e Mice were injected subcutis with agnospheres and, after tumor establishment, randomized in two treatment groups (control and trametinib, 1 mg/Kg/die; AS901,  $n = 7$  mice/group; AS906, control  $n = 6$  mice and trametinib  $n = 7$  mice; AS43,  $n = 6$  mice/group). d Growth curves of tumor volume fold changes vs. day 0 (beginning of treatment; mean  $\pm$  SEM, two-way ANOVA). e Survival curve at the experimental endpoint (tumor volume = 1600 mm<sup>3</sup>; Mantel–Cox test). f Representative H&E and immunohistochemical stainings for Cyclin D1 and phospho-ERK1/2 of whole tumor sections derived from endpoint mice transplanted with AS906 and treated with control or trametinib ( $n = 3$  mice). Scale bar, 2 mm. Bottom row: magnification of pERK1/2 staining. Scale bar, 100  $\mu$ m. g Quantification of necrotic, vital, and Cyclin D1-positive areas in whole tumor sections from mice transplanted with AS906 (top) or AS901 (bottom), expressed as percentage of total tumor area ( $n = 3$  mice/group, mean  $\pm$  SEM, two-way ANOVA, Sidak's multicomparison test; ns not significant). h–j Bar graphs of enrichment scores of the trametinib response signature in agnospheres and melanosphere mMS321 ( $n = 3$  replicates) (h); tumor tissues from CUPs (AGN-), melanomas (MEL-), and early metastatic cancers of known origin (H&N head and neck, BRE breast, MERK merkeloma) (i); and different metastases of CUP patient AGN43 ( $n = 3$  fragments/metastatic site, mean  $\pm$  SEM) (j). Sensitivity to trametinib associates with a negative enrichment score of resistance genes (gene set enrichment analysis). Only enrichment scores with a false discovery rate, FDR < 10% are considered statistically significant,  $p$  values and FDR values are reported in Supplementary Data 1). The presence or absence (WT) of a mutated gene associated with trametinib sensitivity (*BRAF*, *KRAS*, *HRAS*, or *NRAS*) is reported. Hypermuted hypermutated, n/d not determined, ns not significant.

compared with controls, contained an ample central necrotic area surrounded by proliferating cells, still responding to trametinib with ERK1/2 inhibition (Fig. 5f, g and Supplementary Fig. 4j). This indicates that tumor volume measurement alone underestimates the trametinib biological effect, while in vitro proliferation arrest and cell death reliably predict the in vivo tumor tissue response to trametinib.

To attempt an evaluation of trametinib effects on spontaneous metastases, we analyzed serial sections of whole lungs (the site most frequently colonized in spheropatiens), collected at the experimental endpoint, and thus at a later timepoint in treated mice. Nevertheless, we found that, in treated mice, the number and size of metastatic vascular emboli were significantly reduced as compared with controls (Supplementary Fig. 4k, l), indicating that trametinib can also prevent the growth of disseminated cells. To evaluate whether sensitivity to trametinib can be predicted in CUP patients, we set up an original “trametinib response signature”. To this aim, we correlated publically available data on the trametinib response of 445 cancer cell lines (CCLs) with the respective gene expression profiles<sup>42,43</sup> (Supplementary Fig. 4m–o). The signature ability to predict the trametinib response was first validated in an independent dataset of 634 CCLs<sup>44</sup> (Supplementary Fig. 4o), and further confirmed in a panel including the six agnospheres and the *BRAF*-mutated melanosphere mMS321. Here, the signature correctly predicted sensitivity in mMS321, AS901, AS906, AS43, and AS67, and resistance in AS914 and N-AS47 (Fig. 5h), accordingly to in vitro responses (Fig. 5b and Supplementary 4h). The signature was then applied to the transcriptome obtained from paraffin-embedded tissues of a retrospective cohort of patients, encompassing (i) 15 CUPs, including those matched with four agnospheres (Supplementary Tables 1 and 2); (ii) three metastatic *BRAF*-mutated melanomas (including mMEL321 that originated mMS321), chosen as positive controls for trametinib sensitivity; and (iii) ten metastases from early metastatic, non-CUP patients (i.e., enrolled as potential CUPs, but whose tissue of origin was then identified (Supplementary Table 1). The signature predicted trametinib sensitivity in all *BRAF*-mutated melanomas, as expected, and in 11/15 CUPs (73%), while it predicted sensitivity only in 3/10 early metastatic tumors of known origin. Clear resistance was predicted in the neuroendocrine CUP AGN47 (the origin of N-AS47), and in an early metastatic merkeloma (mMERK44), another tumor of neuroectodermal origin. All other cases (3/15 CUPs and 6/10 early metastatic tumors) were below the significance threshold for being defined either as sensitive

or resistant (Fig. 5i). From these data, we can conclude that (i) agnospheres and the corresponding original tumors were predicted consistently, attesting that the signature retains its ability to reliably classify suboptimal samples, such as paraffin-embedded tissues; (ii) the majority of CUP tissues, but only a minor fraction of early metastatic tumors of known origin, associated with trametinib sensitivity independently of the presence of *BRAF* or *RAS* mutations, thus suggesting that constitutive and addictive activation of the MEK signaling pathway can be sustained by alternative mechanisms, including growth factor receptor autocrine loops.

Finally, we also analyzed different synchronous metastases collected at warm autopsy from patient AGN43 (the origin of AS43)<sup>11</sup>, showing that all share the trametinib sensitivity signature (Fig. 5j). This finding adds to the previously shown genetic homogeneity of these metastases<sup>11</sup> and suggests that, at least in some cases, the cell-autonomous behavior of CUP stem-like cells may translate in remarkable biological homogeneity and drug sensitivity across metastatic sites, with favorable implications for CUP therapy.

## Discussion

Isolation of stem-like cells from CUP patients provided, to our knowledge, the first CUP experimental model, suitable to identify genetic and molecular determinants that explain the disease hypermetastatic phenotype, featuring early widespread dissemination, ability to quickly colonize multiple organs by adapting to different microenvironments, and cell differentiation block.

Irrespective of their different genetic alterations, agnospheres display common properties, among which the first and most striking is proliferative and self-renewal autonomy: agnospheres retain their ability to long-term propagate without differentiating even in the absence of any exogenous growth factor, a property seldom displayed by stem-like cells derived from other metastatic tumors. We could associate this property with the cell-autonomous expression of self-renewal and reprogramming TFs, and the constitutive enrichment of a transcriptomic signature distinctive of ES cells<sup>22</sup>. In contrast, in ES cells, this genetic program is driven by extracellular signals, whose withdrawal causes ES differentiation<sup>20</sup>. Within this signature, genes controlled by Polycomb repressors<sup>45</sup> and by the MYC family<sup>46</sup> were strongly modulated, consistently with widespread expression of their transcriptional drivers (BMI1 and EZH2, and MYC, respectively). Interestingly, this ES signature was enriched not

only in agnospheres but also in original CUP tissues, indicating that the entire tumor is mostly formed by cells with stem features, and that, vice versa, agnospheres well represent the overall CUP cell population, thus behaving like “tumoroids”<sup>29</sup>. This also indicates that stem culture conditions used to select agnospheres do not significantly modify the genetic programs expressed by the original CUP cells.

The *c-MYC* proto-oncogene is well known for its role at the crossroad between stem cell regulation and oncogenesis. Famous for being included in the original Yamanaka cocktail that reprograms differentiated fibroblasts to pluripotency<sup>47</sup>, *c-MYC* has been recognized as a key factor to recruit quiescent stem cells into proliferation, by providing not only direct stimulation of the cell cycle, but also activation of metabolic genes that support the bioenergetics needs of proliferating cells<sup>48</sup>. Overexpression of the *c-MYC* proto-oncogene contributes to tumor onset and/or progression, as indicated by findings in human tumors (that overexpress *MYC* in ~30% of cases) and mouse models<sup>38,49</sup>. Experimental *c-MYC* overexpression can confer addiction, so that *MYC* knock-down causes regression of established tumors, attesting its relevance for cell proliferation and survival<sup>50</sup>.

In agnospheres and original CUP tumors, high and constitutive expression of *MYC* family genes (*c-MYC* and/or *N-MYC*) occurs in the absence of *MYC* genetic alterations. *c-MYC* is expressed in the agnospheres derived from the most aggressive cases, diagnosed as “poorly differentiated carcinomas of unknown origin”. *N-MYC* alone is highly expressed in the neuroendocrine CUP, consistently with its specific role in tissues of neuroectodermal origin<sup>51</sup>.

A second feature shared by all agnospheres, which, again, can

be correlated with independency from exogenous signals, is the ability to widely metastasize after subcutaneous transplantation: indeed, agnospheres quickly disseminate, home, survive, and thrive in multiple tissue contexts, where they consistently reproduce the histology of the original tumors, including expression of markers specific to each different patient, together with lack of terminal differentiation. This adaptability is likely conferred by the ability to sustain proliferative and pseudo-differentiative programs in a niche-independent way. Importantly, our data indicate that such prerogative is passed on from patients to the experimental CUP model through agnospheres, likely as result of (epi)genetic mechanisms rendering the proliferative pathway constitutive, either through growth factors autocrine loops or altered signal transduction.

In spite of the above common traits, the agnosphere experimental behavior diverged according to the different clinical courses of the respective original patients: not surprisingly, the frequency of TICs was unusually high and growth of established metastases was rapid in agnospheres isolated from the most aggressive cases, while the same features were decidedly blunted in the N-AS47 from the long-survivor neuroendocrine CUP.

In looking for CUP liabilities, we reasoned that the variegated panel of proliferative signals detected in agnospheres should converge on constitutive activation of the MAPK pathway, known to be a major inducer of *MYC* expression and activity<sup>38,39</sup>. Interestingly, hyperactivation of the MAPK pathway has been recognized as a distinctive feature of metastatic tumors in general<sup>52</sup> and CUP in particular<sup>53</sup>, together with *MYC* overexpression<sup>54</sup>. We could demonstrate that 4/6 agnospheres were addicted to the MAPK pathway and highly sensitive to the specific MEK1/2 inhibitors trametinib or selumetinib<sup>37</sup>. Sensitivity correlated with complete downregulation of *c-MYC* expression, proliferation arrest and cell death in vitro, induction of massive necrosis in experimental tumors growing at the agnosphere injection site, and reduced growth at metastatic sites. Interestingly, in many cancers displaying MAPK hyperactivation, including

those harboring *RAS* mutations, the administration of MEK1/2 inhibitors is poorly effective, as it interrupts negative feedbacks reverberating from the MAPK pathway to tyrosine kinase receptors. These feedbacks may include signals rapidly leading to receptor downregulation, as well as long-term kinome transcriptional reprogramming<sup>37</sup>. Evidence indicates that these mechanisms are likely and consistently disrupted in responsive agnospheres, as trametinib did not induce a rebound effect of tyrosine kinase receptor reactivation, but it rather induced receptor downregulation. Conversely, in non-responsive agnospheres, MEK inhibitors either failed to block cell proliferation, or paradoxically induced it. However, this is not surprising as, for example, in normal ES cells, MEK inhibitors are used to promote rather than to halt proliferation<sup>20</sup>. The mechanisms of resistance to MEK inhibitors remains to be determined, but they likely involve alternative signaling pathways that contribute to maintain high levels of *MYC* gene expression. An attractive candidate is the Wnt pathway, a main regulator of stem cell self-renewal, known to upregulate *MYC*<sup>55</sup>.

To assess whether sensitivity to trametinib, a clinically approved drug, could be predicted in CUP cases, and provide a tool to stratify patients for trials, we elaborated an original “trametinib response signature”. This signature correctly anticipated the experimentally assessed response to trametinib in agnospheres, and was retrieved also in the matched patients’ tissues. Thus validated, the signature predicted the response in a retrospective cohort of CUP cases. Despite the absence of *BRAF* or *RAS* family mutations, usually associated with trametinib sensitivity<sup>37</sup>, the majority of CUPs were classified as responders. Interestingly, CUP sensitivity predicted by the trametinib signature approximates that of *BRAF*-mutated melanomas, while a panel of aggressive metastases from carcinomas of known origin did not display a responder signature. Overall, data presented in this study indicate that constitutive activation of the MAPK pathway, leading to *MYC* sustained expression, and the ensuing stem and proliferative transcriptional programs, can be a CUP prominent, widespread and distinctive pathogenetic mechanism offering opportunities for therapeutic intervention.

Beyond CUP investigation, the integrated experimental platform presented in this study can have far-reaching applications, as it is endowed by unique prerogatives, compared to the metastatic models available so far<sup>34</sup>. First, while such models are mostly based on genetically engineered mice, or the use of conventional cells lines, either artificially manipulated to become metastatic<sup>34</sup>, or injected in the hearth left ventricle to get systemic multi-organ spread<sup>56</sup>, human agnospheres are innately endowed with comprehensive metastatic programs, faithfully inherited from the original tissue and passed on to the mouse model. Moreover, with their properties, agnospheres can help to overcome operational limitations recognized to current metastatic models, such as being (i) slow and inefficient, because generated late in progression by rare cell subpopulations; and (ii) based on cooperation between cell-intrinsic and environmental signals, that hamper successful colonization. The model here described thus represents a next-generation tool for functional validation of metastatic determinants, and mechanistically supported therapeutic interventions in a broad spectrum of aggressive tumors.

## Methods

**Human subjects.** All patients were recruited at the Candiolo Cancer Institute, FPO-IRCCS, according to ethical requirements and protocols approved by the institutional Review Board on human experimentation. Informed consent was obtained from all patients. All patient samples were de-identified before processing. Suspected CUP patients were enrolled in an approved prospective observational trial (Study Protocol N. 010-IRCC-10IIS-15, and following updates; last update: v2.0-16.10.2018), where the diagnosis was attained through an “ad excludendum”



diagnostic protocol in accordance with ESMO guidelines<sup>2</sup>. After diagnosis, patients received the best standard of care, and provided archival and viable tissue specimens, and blood samples.

**Animal models.** All animal procedures were performed according to ethical regulations and protocols approved by the Italian Ministry of Health. NOD.CB17-Prkdc<sup>scid</sup>/NcrCr mice (NOD/SCID), (RRID:IMSR\_CRL:394, Charles River Laboratories), 5- to 6-week-old male were used for all *in vivo* studies. Mice were housed at a maximum of six per cage with a 14-h light/10-h dark cycle, in a conventional animal facility with an ambient temperature and humidity of 20–26 °C and 40%–60%, respectively, with food and water *ad libitum*. Mice were monitored at a minimum of twice weekly for general performance status and euthanized when volume of xenografts reached 1600 mm<sup>3</sup>, or they displayed signs of distress, or weight loss  $\geq$  20%.

**Generation and culture of agnospheres and tumorspheres.** Tumor tissue specimens derived either from CUP patients (AGN906 and AGN47), or PDX (AGN901, AGN43, AGN67, and AGN914), or a melanoma patient (mMEL321) were minced and digested with collagenase I (Gibco) at 2 mg/ml in culture medium for 40 min at 37 °C. After filtration and red cell lysis, single-cell suspensions were resuspended in culture medium, composed by: DMEM:F12 (Sigma), N2 supplement (Life Technologies-GIBCO), BSA 0.5% (Sigma), heparin 4  $\mu$ g/ml (Sigma), 2 mM glutamine (Sigma), penicillin–streptomycin (EuroClone), and seeded in ultra-low-attachment flasks (Corning-Sigma). For AS914 and AS67, the culture medium was supplemented with a chemically defined Lipid Concentrate (Gibco). Culture medium was replaced the next day and used throughout propagation. Spheroids (agnospheres) appeared in culture few days after seeding, and stabilization occurred on average after 3–4 months. For dissociation, trypsin was required for all agnospheres except for AS906 and AS914. mCRC729 and mCRC0155, previously derived from colorectal cancer liver metastases<sup>17</sup>, were kept in the same culture medium as above, supplemented with Lipid Concentrate (Gibco). Agnospheres and tumorspheres will be available upon institutional material transfer agreement approval.

**Generation of patient-derived xenografts.** Human CUP tumor specimens, derived either from biopsy or surgery, were subcutaneously transplanted in NOD/SCID mice (PDX). Tumors were explanted when reached a maximum of 1600 mm<sup>3</sup>, and collected for further propagation and agnosphere derivation.

**Histopathology.** Immunohistochemical staining of formalin-fixed paraffin-embedded (FFPE) tumor tissue sections derived from patients and animal models was performed using Ventana Benchmark ultra System (Roche), or Bond Max (Leica Biosystems) or Autostainer Link 48 (Agilent), according to the antibody used (see reagents), and revealed with Liquid Diaminobenzidine (DAB) + Substrate Chromogen System (K3468, Dako), following standardized manufacturers' instructions. For quantification of experimental tumor areas, stainings of representative middle tumor sections collected at the experimental endpoint were acquired with Nikon Eclipse Ti2 using the NIS Elements Imaging software. The whole area was scanned at 4 $\times$  magnification and analyzed with ImageJ (RRID: SCR\_003070). Percentage of hematoxylin-positive area or Cyclin D1-positive area was normalized vs. total area for each sample. Necrotic areas were calculated as total area minus vital area. Statistical significance was assessed by two-way ANOVA (Sidak's multicomparison test).

**gDNA extraction and CNV analysis.** Genomic DNA was isolated from agnospheres using Relia Prep TM gDNA Tissue Miniprep System (Promega), according to manufacturer's instructions. DNA was quantified using a Nanodrop ND1000 spectrophotometer (Thermo Fisher Scientific). As normal control, gDNA of PBMCs derived from a pool of five healthy individuals was used. *PTEN*, *TP53*, *c-MYC*, and *N-MYC* copy-number variations were calculated after qPCR with the 2<sup>-</sup>1111Ct method using *GREB1* as normalizer.

**Mutational screening of FFPE tissue specimens.** In tumor samples not analyzed by WES, assessment of mutational status of *KRAS*, *HRAS*, *NRAS*, and *BRAF* hotspot mutations was performed through OncoCarta™ Panel v1.0 (Agena Bioscience) on gDNA extracted from FFPE tissues, according to manufacturer's instructions.

**Chromosome G banding.** Chromosome analysis by G banding was performed on agnospheres. In order to increase the number of metaphases, cells were synchronized with Synchronost (Euroclone S.p.A.) according to manufacturer's instructions and blocked with colcemid (10  $\mu$ l/ml) for 1 h. Chromosome harvesting was carried out according to standard procedures. Briefly, cells were incubated in 0.075 M KCl hypotonic solution at 37 °C for 10 min, fixed in methanol–glacial acetic acid (3:1), dropped onto glass slides and dried using specific conditions for optimal chromosome spreading. G banding was performed by incubating slides in 2 $\times$  SSC at 68 °C for 2 min and eventually stained with Wright's solution for 2 min (ref. 57). Metaphase images were captured using an Olympus BX61 microscope (Olympus Corporation)

and analyzed by CytoVision software (Leica Biosystems). An average banding resolution of 300 bands was achieved. Aberrations were described according to the International System for Human Cytogenetic Nomenclature, 2016 (ref. 58).

**Chromosome M-FISH.** Chromosome analysis by multicolor-fluorescence *in situ* hybridization (M-FISH) was carried out on agnospheres by using a commercial mixture containing 24 chromosome-specific painting probes labeled with different fluorochromes (24 $\times$ Cyte kit MetaSystems). Probes and metaphase chromosomes denaturation and hybridization were performed according to manufacturer's instructions. Briefly, the slides were incubated at 70 °C in saline solution (2 $\times$  SSC), denatured in NaOH, dehydrated in an ethanol series, air-dried, then covered with 10  $\mu$ l of the denatured probe cocktail and hybridized for 24 h at 37 °C. Subsequently, the slides were washed with post-hybridization buffers and counterstained with 10  $\mu$ l of DAPI/antifade. The Metafer System and the Metasystems ISIS software (Carl Zeiss) were used for signal detection and metaphase analysis. At least nine metaphases along different culture passages exhibiting the same derivative chromosomes were studied for each cell line.

**RNA-seq and qRT-PCR.** Growing agnospheres and tumorspheres were harvested and RNA was extracted using RNeasy Micro Kit (Qiagen), following manufacturer's instructions. For qRT-PCR mRNA was converted into first-strand cDNA using superscript II Reverse Transcriptase (Invitrogen), according to manufacturer's instructions. Amplification was performed with ABI PRISM 7900 HT (Applied Biosystem) using Taqman Probes (Applied Biosystem). Ct values were normalized versus HPRT1 and represented in the heatmap as 40 – Ct. Results are representative of at least two independent experiments.

For Quant-seq 3' mRNA sequencing, total RNA from agnospheres and tumorspheres or from FFPE patient tissues was purified using RNeasy Micro Kit (Qiagen) or Maxwell® RSC RNA FFPE Kit (Promega), respectively. Total RNA was quantified using the Qubit 2.0 fluorimetric Assay (Thermo Fisher Scientific). Libraries were prepared from 100 ng of total RNA using the QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen GmbH). Quality of libraries was assessed by using screen tape High-sensitivity DNA D1000 (Agilent Technologies). Libraries were sequenced on a NovaSeq 6000 sequencing system using an S1, 100 cycles flow cell (Illumina Inc.).

**Alignment and normalization of QuantSeq RNA data.** Illumina novaSeq base call (BCL) files were converted in fastq file through bcl2fastq (version v2.20.0.422, Illumina Inc.). Sequence reads were trimmed using bbduk software (bbmap suite 37.31, Joint Genome Institute) to remove adapter sequences, poly-A tails, and low-quality end bases (regions with average quality <6). Alignment was performed with STAR 2.6.0a (RRID:SCR\_015899)<sup>59</sup> on hg38 reference assembly obtained from cellRanger website (Single-Cell Gene Expression, 10 $\times$  Genomics; Ensembl Assembly 93). The expression levels of genes were determined with htseq-count 0.9.1 by using cellRanger pre-build genes annotations (Single-Cell Gene Expression, 10 $\times$  Genomics; Ensembl Assembly 93). Genes with an average number of CPM (counts per million) <5 and Perc of duplicated reads >20% were filtered out. Data normalization was performed using edgeR (RRID:SCR\_012802)<sup>60</sup>.

**Analysis of embryonic stem cell signature.** Transcriptional profiles of agnospheres and tumorspheres were first averaged across replicates, and then genes with an average CPM < 1 across all samples removed. Then averaged transcriptional profiles of spheres and transcriptional profiles of CUP tissues were converted in z-score and the gene set enrichment analysis (GSEA)<sup>61</sup> was performed using as input the ES cell signature as described by Ben-Porath et al.<sup>22</sup>. The expression profiles and clinical information of 942 breast cancer patients described by Ben-Porath et al.<sup>22</sup> were downloaded from [barc.wi.mit.edu/benporath](http://barc.wi.mit.edu/benporath). Patients' transcriptional profiles were first converted in z-scores and GSEA performed as above. Then, median enrichment score of each gene signature was plotted grouping patients according to their breast tumor grade. Finally, gene expression profiles of human ES cell lines H7, HUES1, HUES8, and HUES9 were retrieved from ref. 62 with accession number GSE102311, and GSEA was performed after transcriptional profiles were converted in z-scores. GSEA and associated statistics were computed using the fgsea package in R statistical environment version 3.6.

**Generation and validation of the trametinib response signature.** The basal expression profile of ~1000 CCLs was obtained using RNA-seq from Cancer Cell Line Encyclopedia<sup>24,43</sup> (<https://portals.broadinstitute.org/ccle>). Cell lines derived from liquid tumors were discarded, and only cell lines derived from solid tumors were used for the following analysis. The raw counts of each gene across the cancer cell lines were normalized using edgeR<sup>63</sup>. Data on cell line response to trametinib were previously generated by Rees et al.<sup>42</sup> and expressed in terms of area under the curve (AUC), which reflects the *in vitro* response to trametinib of each cancer cell line<sup>24</sup> over 72 h. In particular, lower values of AUC are associated with a higher sensitivity to trametinib and vice versa. A total of 445 CCLs, for which both trametinib response and expression data were available, were used to identify marker genes associated to trametinib resistance, as described below. The gene signature to predict trametinib response from transcriptional data was identified as depicted in Supplementary Fig. 4m–o. First, we computed Pearson correlation

coefficient (PCC) between the expression of each gene and trametinib potency measured by mean of AUC across the 445 distinct CCLs selected, as described above. We then considered the top 1000 genes with highest PCC as putative marker genes of trametinib resistance, because higher expression of these genes is associated with lower potency of trametinib across the panel of the 445 used CCLs. Finally, a machine-learning approach based on recursive feature elimination (RFE) and support vector machines (SVMs) was used to identify the 500 (out of 1000) genes whose expression values were best at discriminating resistant from sensitive trametinib CCLs<sup>64,65</sup>. In particular, we used linear SVMs that were trained and tested using the kernlab package in the R statistical environment version 3.6 (ref. 66). For RFE, trametinib sensitive CCLs were defined as those which AUC values were in the 5% quartile of trametinib AUC distribution across the 445 CCLs used in this study. On the other hand, trametinib resistant CCLs were defined as those whose AUC values were in the 95% quartile. We used GSEA to predict the trametinib response by checking whether marker genes of trametinib resistance identified as above were either downregulated or upregulated across a given transcriptional profile (Supplementary Fig. 4m). Hence, in each transcriptional profile, genes are first sorted from the most to the least expressed and then GSEA is performed using as input the identified marker genes of trametinib resistance. A positive enrichment score means that genes associated with trametinib resistance are highly expressed in the transcriptional profile, thus predicting resistance of these cells to trametinib treatment. Conversely, a negative enrichment score indicates a low expression of marker genes of trametinib resistance, and thus it predicts sensitivity to trametinib treatment (Supplementary Fig. 4n). Validation of the method was performed by predicting the trametinib response in an independent dataset of 634 CCLs from Garnett et al.<sup>44</sup> (Genomics of Drug Sensitivity in Cancer: <https://www.cancerrxgene.org/>), for which basal gene expression from microarray and trametinib response in term of IC50 were available. We then computed the percentage of correctly predicted sensitive and resistant CCLs (Supplementary Fig. 4n). Specifically, CCLs sensitive to trametinib were defined as those with an IC50 < 1  $\mu$ M, while all the other were considered as resistant. We obtained an average classification accuracy of ~76% across the two classes (sensitive and resistant) of CCLs. In all the analyses described above GSEA and associated statistics were computed using the fgsea package in R statistical environment version 3.6.

**Analysis of trametinib response signature.** Agnospheres, tumorspheres, or patient tissues transcriptional profiles were first averaged across replicates when present, then genes with an average CPM < 1 across all samples filtered out. Before applying GSEA, averaged transcriptional profiles of each sample was converted in z-score. GSEA<sup>61</sup> was performed using as input identified marker genes of trametinib resistance with the fgsea package in R statistical environment version 3.6.

**Immunofluorescence and cell immunohistochemistry.** Samples undergoing immunofluorescence were either FFPE tumor specimens or growing agnospheres or colosphere mCRC729. The latter were harvested, fixed 10 min with PFA 4% at 4 °C, washed in PBS, suspended in bio-agar for cyto-inclusion (Bio-Optica) at 42 °C, and processed for inclusion in paraffin. All staining were performed as previously described<sup>67</sup>. Images were acquired using a LEICA SPEII confocal microscope, equipped with a 40 $\times$  oil objective and a 1.5 $\times$  zoom for a final magnification of 600 $\times$ . Optical single sections were acquired with a scanning mode format of 1024  $\times$  1024 pixels. Fluorochrome unmixing was performed by acquisition of automated-sequential collection of multichannel images, to reduce spectral crosstalk between channels. For immunohistochemical staining, an additional peroxidase blocking was performed in H<sub>2</sub>O<sub>2</sub> 3% methanol 50% incubated 20 min in the dark. For primary and secondary antibody concentrations, see reagents. Secondary antibodies were HRP-conjugated (Dako), and DAB substrate chromogen kit (Dako) was used for detection. Nuclei were counterstained with hematoxylin. Images were acquired through LASV4.2 software and are representative of at least three independent immunostainings.

**Western blot analysis.** Total protein were extracted using RIPA buffer supplemented with a protease inhibitor cocktail (Roche Life Science), NaVO<sub>3</sub> 1 mM, and NaF 1 mM, subjected to sonication, quantified using BCA methods (Pierce), and 20  $\mu$ g were separated on SDS-polyacrylamide gradient gel 4–12% or 4–20% (Invitrogen) and blotted onto nitrocellulose membrane. After blocking, primary antibodies were incubated at the indicated concentrations (see reagents). After incubation with HRP-conjugated secondary antibodies (Jackson Lab), enhanced chemiluminescence (Biorad) was used for detection according to manufacturer's instructions and images were acquired with the ChemiDoc Touch<sup>TM</sup> Imaging System (Biorad) through Image Lab software.  $\beta$ -Actin or  $\beta$ 2-tubulin were used as protein loading control as indicated. The results shown are representative of at least three independent experiments.

**In vitro limiting dilution assay.** Agnospheres were dissociated and seeded at limiting dilution concentration (1–100 cells/100  $\mu$ l) in ultra-low-attachment 96-well microtiter (Corning). "Positive tests" were defined as wells with primary spheres with a diameter  $\geq$  100  $\mu$ m. Stem cell frequency was calculated using the ELDA software<sup>68</sup> (<http://bioinf.wehi.edu.au/software/elda/>). Means and 95%

confidence intervals (CI) are shown ( $n \geq 3$  independent experiments). Primary spheres were harvested, dissociated, and seeded at the same dilutions for a second and a third passage to assess long-term propagation.

**Cell viability.** Agnospheres were dissociated and seeded in culture medium in 96-well microtiter wells at the concentration of 200 cells/100  $\mu$ l in the case of AS901, AS906, AS43, and AS67, 500 cells/100  $\mu$ l in the case of AS914 or 1000 cells/100  $\mu$ l in the case of N-AS47. Trametinib (GSK1120212; Selleckchem) or vehicle (DMSO) were added immediately after seeding and every 5 days at the indicated concentrations. ATP consumption was measured 4 or 10 days after cell seeding with Cell Titer Glo (Promega) according to manufacturer's instructions, using GloMax 96 Microplate Luminometer and GloMax<sup>®</sup> 96 software (Promega). In each experiment, the average of relative luminescence values ( $n \geq 6$  technical replicates) was normalized vs. untreated control and fold changes were reported ( $n \geq 3$  mean  $\pm$  SEM, ANOVA, Bonferroni multicomparison test).

**In vivo agnosphere transplantation.** To assess the tumorigenic potential, agnospheres were dissociated as single-cell suspensions and counted with trypan blue to exclude dead cells. A total of 10<sup>5</sup> cells were resuspended in 50  $\mu$ l of culture medium mixed 1:1 with matrigel—growth factor reduced (BD Bioscience) and injected subcutaneously in the flank of 5–6-week-old male NOD.CB17-Prkdcscid/J mice (referred to as "spheropatient"), obtaining 100% engraftment efficiency. During the procedure anesthesia with 2.5% isoflurane in 100% oxygen at a flow rate of 1 l/min was delivered to mice. Tumors grown at the injection site (IS-tumors) were explanted and FFPE to undergo histopathological evaluation and comparison with original tumors, as detailed above for patients.

**Tumor-initiating cell frequency (in vivo limiting dilution assay).** For in vivo LDA, single-cell suspensions of 10<sup>5</sup>, 10<sup>4</sup>, 10<sup>3</sup>, 10<sup>2</sup>, or 10 cells were subcutaneously injected into immunocompromised mice (P0) as above. Tumor intake (formation of IS-tumors) was monitored up to 7 months after transplantation or until spontaneous death, and the in vivo TIC frequency was calculated with the ELDA software as above. Mice dead before the experimental endpoint that did not generate a tumor were excluded from the counts. From P0 tumors agnospheres were re-derived in culture by enzymatic digestion with collagenase as above, and, after 1 week of recovery in culture, they were dissociated and transplanted for a second passage in mice (P1) at the same cell dilutions as above<sup>69</sup>. In some experiments, the procedure was further repeated for a third passage (P2).

**Assessment of spontaneous metastases.** Agnospheres and colosphere mCRC729 were engineered to express luciferase. Briefly, single-cell suspensions were seeded in ultra-low-attachment six-well microplates at 5  $\times$  10<sup>5</sup> cells/well concentration in 1.5 ml of culture medium. Lentiviral particles carrying pCMV-Luciferase or pCMV-Luciferase-IRES-GFP transfer vectors were added at 5 MOI. After 6 h, 1 ml medium was added, and 24 h later the medium was replaced. Transduction efficiency was assessed at 72 h by fluorescence microscopy showing 98% in all agnospheres. Transduced agnospheres were dissociated to single-cell suspensions, and 5  $\times$  10<sup>4</sup> cells were injected as detailed above in the flank of NOD/SCID mice. Within a month after transplantation, tumors reached ~300 mm<sup>3</sup> and were surgically removed under anesthesia (as above) to prevent tumor overgrowth during subsequent metastasis monitoring. Tumor volume was measured with a caliper and calculated using the formula:  $(d^2 \times D)/2$ , where  $d$  and  $D$  are the minor and the major tumor axis, respectively. Analgesics were delivered to the animals according to institutional guidelines. Bioluminescence in vivo imaging was acquired with IVIS<sup>®</sup> Lumina and IVIS imaging software (Caliper Life Sciences) starting from the day after surgery. Luciferin (D-Luciferin potassium salt, Caliper Life Sciences) dissolved in PBS (150 mg/kg) was administered to mice by subcutaneous injection. Anesthesia was delivered in the induction chamber with 2.5% isoflurane in 100% oxygen at a flow rate of 1 l/min and maintained during IVIS imaging with a 1.5% isoflurane mixture as above at 0.5 l/min. To mask residual signals from the subcutaneous tumor, a black tape was applied during imaging. Luminescent signals were monitored weekly until mice reached the clinical endpoint. For ex vivo imaging, luciferin was administered as above, mice were euthanized by carbon dioxide inhalation and all organs were immediately explanted and analyzed with IVIS. Organs were FFPE to undergo histopathological evaluation and comparison with original tumors as above. The frequency of metastatic sites was assessed by ex vivo organ luminescence and by immunohistochemistry of CUP-specific markers in organ sections. Such frequency could be underestimated for missed identification of micrometastases in some organs. To assess early dissemination, mice transplanted with AS43 were euthanized 10 days after subcutaneous injection, and explanted organs were FFPE and analyzed as above.

**In vivo trametinib treatment.** Unlabeled agnospheres were inoculated subcutis in immunocompromised mice. When tumors reached ~150 mm<sup>3</sup>, mice were randomized in two groups (control and trametinib-treated) using LAS software<sup>70</sup>. For treatment, trametinib was dissolved in hydroxymethylcellulose 0.5%+Tween 80 and administered by oral gavage at 1 mg/Kg/die (therapeutic range in preclinical models: 0.3–3 mg/kg/die)<sup>41</sup>. Tumor volume was measured and calculated as above,



and expressed as fold changes vs. day 0 for each mouse. For each experiment,  $n \geq 6$  mice/group were used. Statistical significance was determined by two-way ANOVA. The survival curve was generated considering a tumor volume of 1600 mm<sup>3</sup> as experimental endpoint, and statistical significance was calculated using log-rank (Mantel–Cox test).

**Statistical analysis.** Results were expressed as mean  $\pm$  standard error of the mean (SEM) of at least three independent experiments. Statistical comparisons were performed using Prism v8.0 software (RRID:SCR\_005375; GraphPad). Chi-squared test was used for limiting dilution experiments performed by ELDA software.

ANOVA (Bonferroni's multiple comparison test correction) was used for assessment of in vitro trametinib and selumetinib effect. Two-way ANOVA was used for tumor growth curves, Mantel–Cox test for survival curves and two-way ANOVA (Sidak's for multiple comparisons) for evaluation of vital/necrotic tumor areas. A  $p$  value  $< 0.05$  or a false discovery rate  $< 10\%$  was considered significant. For RNA-seq analysis the *fgsea* package in R statistical environment version 3.6 was used.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Complete datasets related to WES analysis of somatic mutations, referring to Supplementary Table 3, were deposited in the European Genome-phenome Archive (EGA), under the accession code [EGAS00001004868](https://ega-archive.org/studies/EGAS00001004868). Raw data relative to 3'UTR-Seq, referring to Figs. 2d and 5g–I and Supplementary Fig. 1d, e were uploaded in the Gene Expression Omnibus (GEO) repository, under the accession code [GSE167473](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE167473). Datasets from public databases are available at these web links: Cancer Cell Line Encyclopedia: <https://portals.broadinstitute.org/ccle>; Genomics of Drug Sensitivity in Cancer: <https://www.cancerxgene.org/>. Source data are provided with this paper.

Received: 19 July 2020; Accepted: 16 March 2021;

Published online: 03 May 2021

## References

- Varadhachary, G. R. & Raber, M. N. Cancer of unknown primary site. *N. Engl. J. Med.* **371**, 757–765 (2014).
- Fizazi, K. et al. Cancers of unknown primary site: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **26**(Suppl 5), v133–138 (2015).
- Kolling, S. et al. "Metastatic cancer of unknown primary" or "primary metastatic cancer"? *Front. Oncol.* **9**, 1546 (2019).
- Rassy, E. & Pavlidis, N. Progress in refining the clinical management of cancer of unknown primary in the molecular era. *Nat. Rev. Clin. Oncol.* **17**, 541–554 (2020).
- Moran, S., Martínez-Cardús, A., Boussios, S. & Esteller, M. Precision medicine based on epigenomics: the paradigm of carcinoma of unknown primary. *Nat. Rev. Clin. Oncol.* **14**, 682–694 (2017).
- Moran, S. et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol.* **17**, 1386–1395 (2016).
- Tohill, R. W. et al. Development and validation of a gene expression tumour classifier for cancer of unknown primary. *Pathology* **47**, 7–12 (2015).
- Ross, J. S. et al. Comprehensive genomic profiling of carcinoma of unknown primary site: new routes to targeted therapies. *JAMA Oncol.* **1**, 40–49 (2015).
- Varghese, A. M. et al. Clinical and molecular characterization of patients with cancer of unknown primary in the modern era. *Ann. Oncol.* **28**, 3015–3021 (2017).
- Hayashi, H. et al. Randomized phase II trial comparing site-specific treatment based on gene expression profiling with carboplatin and paclitaxel for patients with cancer of unknown primary site. *J. Clin. Oncol.* **37**, 570–579 (2019).
- Benvenuti, S. et al. Cancer of unknown primary (CUP): genetic evidence for a novel nosological entity? A case report. *EMBO Mol. Med.* **12**, e11756 (2020).
- Pastrana, E., Silva-Vargas, V. & Doetsch, F. Eyes wide open: a critical review of sphere-formation as an assay for stem cells. *Cell Stem Cell* **8**, 486–498 (2011).
- Pece, S. et al. Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell* **140**, 62–73 (2010).
- Galli, R. et al. Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res.* **64**, 7011–7021 (2004).
- Luraghi, P. et al. MET signaling in colon cancer stem-like cells blunts the therapeutic response to EGFR inhibitors. *Cancer Res.* **74**, 1857–1869 (2014).
- De Bacco, F. et al. The MET oncogene is a functional marker of a glioblastoma stem cell subtype. *Cancer Res.* **72**, 4537–4550 (2012).
- Luraghi, P. et al. A molecularly annotated model of patient-derived colon cancer stem-like cells to assess genetic and nongenetic mechanisms of resistance to anti-EGFR therapy. *Clin. Cancer Res.* **24**, 807–820 (2018).
- Wainwright, E. N. & Scaffidi, P. Epigenetics and cancer stem cells: unleashing, hijacking, and restricting cellular plasticity. *Trends Cancer* **3**, 372–386 (2017).
- De Los Angeles, A. et al. Hallmarks of pluripotency. *Nature* **525**, 469–478 (2015).
- Martello, G. & Smith, A. The nature of embryonic stem cells. *Annu. Rev. Cell Dev. Biol.* **30**, 647–675 (2014).
- Pancieria, T. et al. Induction of expandable tissue-specific stem/progenitor cells through transient expression of YAP/TAZ. *Cell Stem Cell* **19**, 725–737 (2016).
- Ben-Porath, I. et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.* **40**, 499–507 (2008).
- Shibue, T. & Weinberg, R. A. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat. Rev. Clin. Oncol.* **14**, 611–629 (2017).
- Stemmler, M. P., Eccles, R. L., Brabletz, S. & Brabletz, T. Non-redundant functions of EMT transcription factors. *Nat. Cell Biol.* **21**, 102–112 (2019).
- Medema, J. P. Cancer stem cells: the challenges ahead. *Nat. Cell Biol.* **15**, 338–344 (2013).
- Casey, S. C. et al. MYC regulates the antitumor immune response through CD47 and PD-L1. *Science* **352**, 227–231 (2016).
- Feng, M. et al. Phagocytosis checkpoints as new targets for cancer immunotherapy. *Nat. Rev. Cancer* **19**, 568–586 (2019).
- Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14**, 275–291 (2014).
- Tuveson, D. & Clevers, H. Cancer modeling meets human organoid technology. *Science* **364**, 952–955 (2019).
- Ishizawa, K. et al. Tumor-initiating cells are rare in many human tumors. *Cell Stem Cell* **7**, 279–282 (2010).
- Kreso, A. et al. Self-renewal as a therapeutic target in human colorectal cancer. *Nat. Med.* **20**, 29–36 (2014).
- Bartucci, M. et al. TAZ is required for metastatic activity and chemoresistance of breast cancer stem cells. *Oncogene* **34**, 681–690 (2015).
- Richichi, C. et al. Tumor-initiating cell frequency is relevant for glioblastoma aggressiveness. *Oncotarget* **7**, 71491–71503 (2016).
- Gómez-Cuadrado, L., Tracey, N., Ma, R., Qian, B. & Brunton, V. G. Mouse models of metastasis: progress and prospects. *Dis. Model Mech.* **10**, 1061–1074 (2017).
- Yarden, Y. & Pines, G. The ERBB network: at last, cancer therapy meets systems biology. *Nat. Rev. Cancer* **12**, 553–563 (2012).
- Lito, P. et al. Disruption of CRAF-mediated MEK activation is required for effective MEK inhibition in KRAS mutant tumors. *Cancer Cell* **25**, 697–710 (2014).
- Caunt, C. J., Sale, M. J., Smith, P. D. & Cook, S. J. MEK1 and MEK2 inhibitors and cancer therapy: the long and winding road. *Nat. Rev. Cancer* **15**, 577–592 (2015).
- Meyer, N. & Penn, L. Z. Reflecting on 25 years with MYC. *Nat. Rev. Cancer* **8**, 976–990 (2008).
- Bretones, G., Delgado, M. D. & León, J. Myc and cell cycle control. *Biochim. Biophys. Acta* **1849**, 506–516 (2015).
- Kawasaki, K. et al. An organoid biobank of neuroendocrine neoplasms enables genotype-phenotype mapping. *Cell* **183**, 1420–1435 (2020).
- Gilmartin, A. G. et al. GSK1120212 (JTP-74057) is an inhibitor of MEK activity and activation with favorable pharmacokinetic properties for sustained in vivo pathway inhibition. *Clin. Cancer Res.* **17**, 989–1000 (2011).
- Rees, M. G. et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* **12**, 109–116 (2016).
- Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
- Garnett, M. J. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
- Lee, T. I. et al. Control of developmental regulators by polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
- Fernandez, P. C. et al. Genomic targets of the human c-Myc protein. *Genes Dev.* **17**, 1115–1129 (2003).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Stine, Z. E., Walton, Z. E., Altman, B. J., Hsieh, A. L. & Dang, C. V. MYC, metabolism, and cancer. *Cancer Discov.* **5**, 1024–1039 (2015).
- Dang, C. V. MYC on the path to cancer. *Cell* **149**, 22–35 (2012).
- Arvanitis, C. & Felsher, D. W. Conditional transgenic models define how MYC initiates and maintains tumorigenesis. *Semin. Cancer Biol.* **16**, 313–317 (2006).

51. Stanton, B. R. & Parada, L. F. The N-myc proto-oncogene: developmental expression and in vivo site-directed mutagenesis. *Brain Pathol.* 2, 71–83 (1992).
52. Khan, I. & Steeg, P. S. Metastasis suppressors: functional pathways. *Lab. Invest.* 98, 198–210 (2018).
53. Krikelis, D. et al. Profiling immunohistochemical expression of NOTCH1-3, JAGGED1, cMET, and phospho-MAPK in 100 carcinomas of unknown primary. *Clin. Exp. Metastasis* 29, 603–614 (2012).
54. Pavlidis, N., Briassoulis, E., Bai, M., Fountzilas, G. & Agnantis, N. Overexpression of C-myc, Ras and C-erbB-2 oncoproteins in carcinoma of unknown primary origin. *Anticancer Res.* 15, 2563–2567 (1995).
55. He, T. C. et al. Identification of c-MYC as a target of the APC pathway. *Science* 281, 1509–1512 (1998).
56. Jin, X. et al. A metastasis map of human cancer cell lines. *Nature* 588, 331–336 (2020).
57. Cavalloni, G. et al. Establishment and characterization of a human intrahepatic cholangiocarcinoma cell line derived from an Italian patient. *Tumour Biol.* 37, 4041–4052 (2016).
58. McGowan-Jordan, J., Simons, A. & Schmid, M. *ISCN2016: An International System for Human Cytogenomic Nomenclature* (Karger, 2016).
59. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013).
60. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169 (2015).
61. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* 102, 15545–15550 (2005).
62. Sun, C. et al. Transcriptome variations among human embryonic stem cell lines are associated with their differentiation propensity. *PLoS ONE* 13, e0192625 (2018).
63. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2010).
64. Gambardella, G. et al. The impact of microRNAs on transcriptional heterogeneity and gene co-expression across single embryonic stem cells. *Nat. Commun.* 8, 14126 (2017).
65. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422 (2002).
66. Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A. kernlab-AnS4P package for kernel methods in R. *J. Stat. Softw.* 11, 1–20 (2004).
67. Verginelli, F. et al. Transcription factors FOXG1 and Groucho/TLE promote glioblastoma growth. *Nat. Commun.* 4, 2956 (2013).
68. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* 347, 70–78 (2009).
69. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* 445, 106–110 (2007).
70. Baralis, E., Bertotti, A., Fiori, A. & Grand, A. LAS: a software platform to support oncological data management. *J. Med. Syst.* 36(Suppl 1), S81–S90 (2012).

## Acknowledgements

We thank P. Luraghi for mCRC729 and mCRC0155 generation, S. Kolling, F. Ventre, and A. Polidori for help with patients' data, the TIGEM Bioinformatics Core for help

with bioinformatic analysis, I. Sarotto, M. Milan, P. Ferrero, R. Albano, S. Giove, and R. Carollo for technical help, the TIGEM NGS facility for sequencing, C. Marchiò for organizational help, A. Balsamo, A. Cignetto, and D. Gramaglia for assistance. This work was supported by AIRC—Italian Association for Cancer Research (“Special Program Molecular Clinical Oncology 5 × 1000”, N. 9970 and N. 21052, and Investigator Grant N. 19933 to C.B.); FPRC 5 × 1000 “2014” and RC “2019”, Ministero della Salute; F.V. and E. Candiello were recipients of “Fondazione U. Veronesi” postdoctoral fellowships. G.G. was supported by the STAR (Sostegno Territoriale alle Attività di Ricerca) grant of University of Naples Federico II and My First AIRC Grant 23162.

## Author contributions

F.V. and C.B. conceived the project, planned experiments, interpreted data, and wrote the manuscript with inputs from all the authors; F.V. was involved in performing all the experiments; G.G. performed bioinformatic analysis; A.P., E. Cascardi, A.S., and R.S. performed histopathological diagnosis in patients and models; A.D.A. helped with in vivo, and E. Candiello and M.F. with in vitro experiments; M.P. and L.C. analyzed cytogenetics; S.B. contributed to genetic analysis; E.G. and F.M. managed patients; A.B. supervised bioinformatic analysis; and P.M.C. and C.B. supervised the project.

## Competing interests

A.B. is a cofounder of CASMA Therapeutics Inc. All other authors declare that they have no competing interests.

## Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-22643-w>.

Correspondence and requests for materials should be addressed to C.B.

Peer review information *Nature Communications* thanks Alwin Krämer, Rina Plattner and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

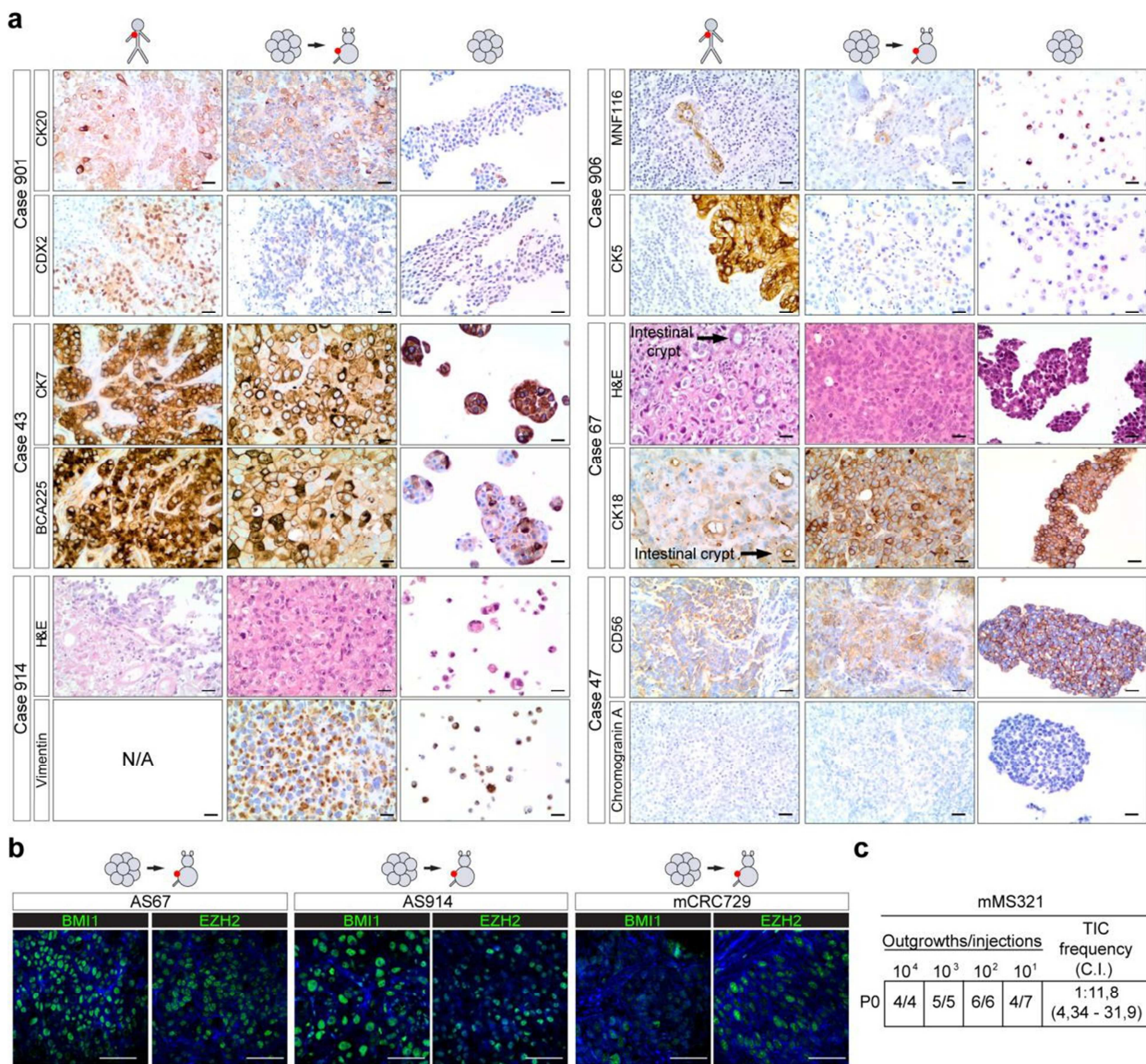


Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

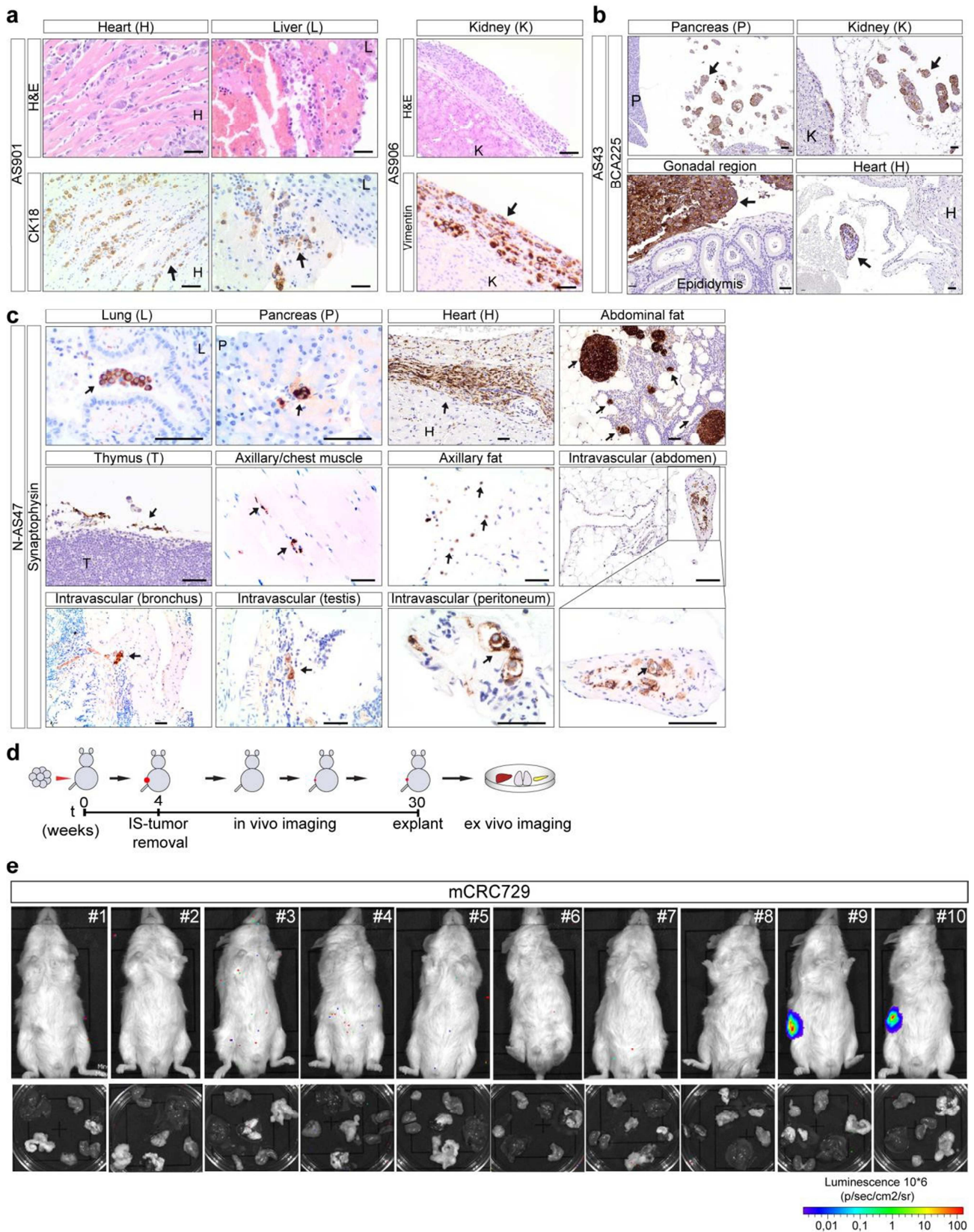




**Supplementary Figure 2.** Agnospheres generate phenocopies of the original tumors. **a**, Histopathological features (H&E) and immunohistochemistry with antibodies used for CUP diagnosis, in: patients' metastases, tumors formed at the subcutaneous injection site (IS-tumor) and agnospheres. CDX2 and CK5 are expressed only in focal areas of original tumors AGN901 and AGN906, respectively (see Supplementary Table 1). A representative image of IS-tumor is shown (n=3 independent stainings, for each marker similar results were obtained). The patient tumor tissue of case 67 includes intestinal crypts (arrows) as it was excised from a mass infiltrating the colonic wall, carefully excluded to be a primary site. Scale bar, 50  $\mu$ m. **b**, Immunofluorescent staining of stem

cell markers in tumors formed at the subcutaneous injection site of agnospheres AS67 and AS914 and metastatic colosphere mCRC729. A representative image of IS-tumor is shown (n=3 independent stainings, for each marker similar results were obtained). Scale bar, 50  $\mu$ m. BCA225: breast cancer antigen 225; CDX2: caudal type homeobox 2; CK: cytokeratin; H&E: hematoxylin and eosin; N/A: not available. **c**, *In vivo* limiting dilution assay of melanosphere mMS321. The indicated numbers of melanosphere cells ( $10^4$ - $10^1$ ) were transplanted *subcutis* into immunocompromised mice (P0). TIC: Tumor-initiating cell frequency. C.I.: confidence interval.





**Supplementary Figure 3.** Subcutaneously transplanted agnospheres reproduce the early and multi-organ metastatic pattern of CUP patients while spheres from metastatic

colorectal cancer do not metastasize. **a-c**, Histopathology (H&E staining) and immunohistochemistry of metastases in the indicated organs from spheropatiens, performed with human-specific antibodies against markers ubiquitously expressed in the original CUPs. Representative images are shown for each organ. **a**, AS901 spheropatient, immunostaining for Cytokeratin 18 (CK18); AS906 spheropatient, immunostaining for Vimentin. Scale bar, 50  $\mu$ m. **b**, AS43 spheropatient, immunostaining for BCA225. Scale bar, 50  $\mu$ m. **c**, N-AS47 spheropatient, immunostaining for Synaptophysin. Inset: image magnification is shown in the row below. Scale bar, 50  $\mu$ m. **d**, Outline of longitudinal *in vivo* monitoring of spheropatiens transplanted *subcutis* with luciferase-labelled colosphere mCRC729, followed by end-point *ex-vivo* organ imaging with IVIS. IS-tumor: injection site tumor. **e**, Representative *in vivo* images of mice transplanted with mCRC729 (n=14 mice) and *ex-vivo* images of different organs (bottom), taken 30 weeks after transplantation, showing absence of bioluminescent signals suggestive of metastases. Signals detected in mice #9 and #10 (top) correspond to local regrowth after IS-tumor removal.



SOURCE  
DATATRANSPARENT  
PROCESSOPEN  
ACCESS

# Mutated axon guidance gene *PLXNB2* sustains growth and invasiveness of stem cells isolated from cancers of unknown primary

Serena Brundu<sup>1,†</sup> , Virginia Napolitano<sup>2,†</sup> , Giulia Franzolin<sup>1,†</sup> , Ettore Lo Cascio<sup>3</sup> , Roberta Mastrantonio<sup>2</sup> , Gabriele Sardo<sup>1</sup>, Eliano Cascardi<sup>1,4</sup> , Federica Verginelli<sup>1</sup> , Sergio Sarnataro<sup>5</sup> , Gennaro Gambardella<sup>5,6</sup> , Alberto Pisacane<sup>1</sup>, Alessandro Arcovito<sup>3,7</sup> , Carla Boccaccio<sup>1,8</sup> , Paolo M Comoglio<sup>9</sup> , Enrico Giraud<sup>1,10,\*;‡</sup> & Luca Tamagnone<sup>2,7,\*;‡</sup>

## Abstract

The genetic changes sustaining the development of cancers of unknown primary (CUP) remain elusive. The whole-exome genomic profiling of 14 rigorously selected CUP samples did not reveal specific recurring mutation in known driver genes. However, by comparing the mutational landscape of CUPs with that of most other human tumor types, it emerged a consistent enrichment of changes in genes belonging to the axon guidance KEGG pathway. In particular, G842C mutation of PlexinB2 (PlxnB2) was predicted to be activating. Indeed, knocking down the mutated, but not the wild-type, PlxnB2 in CUP stem cells resulted in the impairment of self-renewal and proliferation in culture, as well as tumorigenic capacity in mice. Conversely, the genetic transfer of G842C-PlxnB2 was sufficient to promote CUP stem cell proliferation and tumorigenesis in mice. Notably, G842C-PlxnB2 expression in CUP cells was associated with basal EGFR phosphorylation, and EGFR blockade impaired the viability of CUP cells reliant on the mutated receptor. Moreover, the mutated PlxnB2 elicited CUP cell invasiveness, blocked by EGFR inhibitor treatment. In sum, we found that a novel activating mutation of the axon guidance gene *PLXNB2* sustains proliferative autonomy and confers invasive properties to stem cells isolated from cancers of unknown primary, in EGFR-dependent manner.

Keywords CUP; EGFR; exome; mutation; Plexin

Subject Categories Cancer; Chromatin, Transcription & Genomics; Molecular Biology of Disease

DOI 10.15252/emmm.202216104 | Received 31 March 2022 | Revised 28

December 2022 | Accepted 11 January 2023 | Published online 1 February 2023

EMBO Mol Med (2023) 15: e16104

## Introduction

Cancer of unknown primary (CUP) is a heterogeneous clinical syndrome and pathological entity represented by metastatic tumors that are first time diagnosed in the absence of a clinically detectable primary lesion (Fizazi *et al*, 2015; Rassy & Pavlidis, 2020). CUPs account for about 2–5% of all cancer diagnoses and are characterized by a rapidly progressing clinical course and short median survival (< 1 year) (Pavlidis & Pentheroudakis, 2012; Olivier *et al*, 2021). Thus far, although high throughput sequencing has been applied to CUP samples, their genetic landscape remains poorly understood, and the molecular mechanisms underpinning the early metastatic dissemination still await elucidation. In most recent years, CUPs were subjected to genomic surveys focused on panels of genes considered oncogenic drivers and selected according to their relevance as potential therapeutic targets (Ross *et al*, 2015, 2021; Löffler *et al*, 2016; Subbiah *et al*, 2017; Varghese *et al*, 2017; Zehir *et al*, 2017; Benvenuti *et al*, 2020).

1 Candiolo Cancer Institute, FPO-IRCCS, Turin, Italy

2 Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Rome, Italy

3 Department of Biotechnological Sciences and Intensive Care, Università Cattolica del Sacro Cuore, Rome, Italy

4 Department of Medical Sciences, University of Turin, Turin, Italy

5 Telethon Institute of Genetic and Medicine, Pozzuoli, Italy

6 Department of Electrical Engineering and Information Technology, University of Naples Federico II, Naples, Italy

7 Fondazione Policlinico Gemelli (FPG) – IRCCS, Rome, Italy

8 Department of Oncology, University of Turin, Turin, Italy

9 IFOM, FIRC Institute of Molecular Oncology, Milan, Italy

10 Department of Science and Drug Technology, University of Turin, Turin, Italy

\*Corresponding author. Tel: +39 011 9933279; E-mail: enrico.giraud@ircc.it

\*\*Corresponding author. Tel: +39 063 0154463; E-mail: luca.tamagnone@unicatt.it

†These authors contributed equally to this work

‡These authors contributed equally to this work as co-senior authors

In all studies, the most recurrent and biologically relevant alterations observed were those affecting TP53 (about 50% of the samples), while CUP-specific genetic changes did not emerge.

Notably, beyond the impact of individual mutations, the genetic landscape of other cancer types revealed the enrichment of diverse alterations in components of the same functional pathway. For example, genomic analyses revealed changes in axon guidance pathway genes in highly metastatic tumors such as pancreatic adenocarcinoma (Biankin *et al*, 2012) and neuroblastoma (Li *et al*, 2017). Indeed, axon guidance cues, beyond their role in the wiring of the neural network, are known to control a wide range of cell migration processes and are increasingly implicated in human cancer (Ch.edotal *et al*, 2005). For example, neural crest cells disseminate throughout the organism in response to specific attractive and repelling cues encoded by genes in this family (e.g., SDF1/CXCL12, Neuregulin, GDNF/Artemin, and Sema3A/3F). Other signals controlling both axonal navigation and cell migration include hepatocyte growth factor, and large families of guidance cues conserved in evolution: Semaphorins, Ephrins, Netrins, and Slits (Hinck, 2004). In fact, the specific receptors for these signals are expressed on the surface of migrating cells and extending axons. In particular, semaphorins form the largest family of guidance cues, either secreted or membrane bound, and their specific transmembrane receptors are represented by the Plexins, counting nine members in humans (Yazdani & Terman, 2006; Neufeld *et al*, 2016). This complex and versatile vocabulary of signaling molecules is widely used in development, while most guidance cues are downregulated in the adult stage; yet, their functional relevance resumes in case of tissue remodeling and regeneration, as well as in cancer (Ch.edotal *et al*, 2005; Fard & Tamagnone, 2021).

In the present work, we aimed at characterizing the genetic landscape of the CUP syndrome, undertaking for the first time a whole-exome analysis on a cohort of unambiguously selected CUP patients, in order to discover any genetic dependencies and functional vulnerabilities. In addition, we explored the potential enrichment in CUPs, of mutations affecting specific signaling pathways; this revealed a putative functional role of axon guidance genes, and of the mutated semaphorin receptor *PLXNB2* in particular. Notably, previous studies have implicated this axon guidance receptor in the regulation of cancer cell proliferation, invasiveness, and metastatic spreading (Le *et al*, 2015; Yu *et al*, 2017; Gurrapu *et al*, 2018; Huang *et al*, 2021); however, nothing was known about *PLXNB2* mutations in cancer, or in other settings.

In order to study the functional role of mutated *PLXNB2* in CUP development, we exploited a validated experimental model in culture, that is, cancer stem cell-enriched spheroids derived from CUP biopsies (Verginelli *et al*, 2021), which have been previously demonstrated to faithfully recapitulate the original CUP phenotype and genetic landscape, including *PLXNB2* mutation. Our data indicate that G842C-*PLXNB2* is a novel genetic change enhancing CUP stem cell proliferation, tumorigenic capacity, and EGFR kinase-dependent invasiveness, providing a bona fide proof-of-principle of the functional involvement of mutated axon guidance genes in CUP development.

## Results

### Whole-exome sequencing of metastasis in a cohort of CUP patients

We undertook a comprehensive genetic analysis of metastases biopsied from 14 CUP patients, unambiguously selected adopting the 2015 ESMO guidelines (Fizazi *et al*, 2015), as described in the [Materials and Methods](#) section. Patients included seven women and seven men, with a median age at presentation of 57 (range 42–71) years; one patient had a single metastasis, while the other 13 presented with multiple metastases at diagnosis. The clinicopathological details of these cases are listed in Table EV1 (while Appendix Fig S1 shows representative histological images).

We performed whole-exome sequencing (WES) on DNA extracted from this collection of CUP samples; gDNA from respective patients' PBMCs was used as normal matched control. Genetic alterations have been detected in all samples. However, it was possible to identify three subgroups in the cohort: five tumors may be defined as hypomutated (displaying < 50 genetic alterations per exome), six were normomutated (displaying from 50–250 mutations), and three were hypermutated (with more than 10<sup>6</sup> mutations per sequenced megabase; Campbell *et al*, 2017); the hypermutated status was associated with mutations in either *BRCA1*, *BRCA2*, or *POLE* (see Table EV1).

When ranking genetic changes, based on the frequency in the whole panel of CUPs, the top mutated gene *TP53* was altered in 43% of the samples, consistent with previous evidence in the literature (Tothill *et al*, 2013; Ross *et al*, 2015). The transmembrane transporter *ABCA13*, the Sushi Domain-Containing Protein 3 precursor *CSMD3*, the two proteins *DNAH2* and *DNAH9* involved in ATP hydrolysis and flagella movement, the axon guidance cue *PLXNA4*, the extracellular matrix protein *USH2A*, and the two transcription factors *ZFH3* and *ZFH4* were mutated in 36% of the analyzed CUP samples. We furthermore expectedly found frequent mutations in giant genes *TTN* (57% of samples) and *OBSCN* (mutated in 36%), which are at high risk of genetic changes due to random DNA repair errors (Tan *et al*, 2015).

We wondered whether any functional pathways were more frequently hit by genetic changes in CUPs. To answer this question, we applied MEGA-V algorithm (Gambardella *et al*, 2017) to comparatively analyze the status of genes grouped in 186 KEGG pathways (Kanehisa & Goto, 2000), within our panel of CUP samples versus in datasets derived from 33 different cohorts of patients affected by other tumor types (TCGA mutation database; <https://gdc.cancer.gov/>). Notably, in comparison with nine tumor types with higher mutational load, CUPs did not show a significantly greater number of genetic changes in any of the KEGG functional pathways. However, based on 24 informative tumor comparisons, certain pathways arose as more frequently mutated in CUPs in statistically significant manner (FDR < 0.1; see Dataset EV1). Among top hits, the Axon Guidance pathway (hsa04360; <https://www.genome.jp/entry/pathway+hsa04360>) caught our attention, due to its emerging role in cancer progression and metastasis (Ch.edotal *et al*, 2005), and because it included one of the top frequently mutated genes in our CUP cohort, *PLXNA4*. In addition, the axon guidance gene *MET* had been found mutated in a panel of stringently selected CUP cases (Stella *et al*, 2011). Importantly, MEGA-V analysis showed that

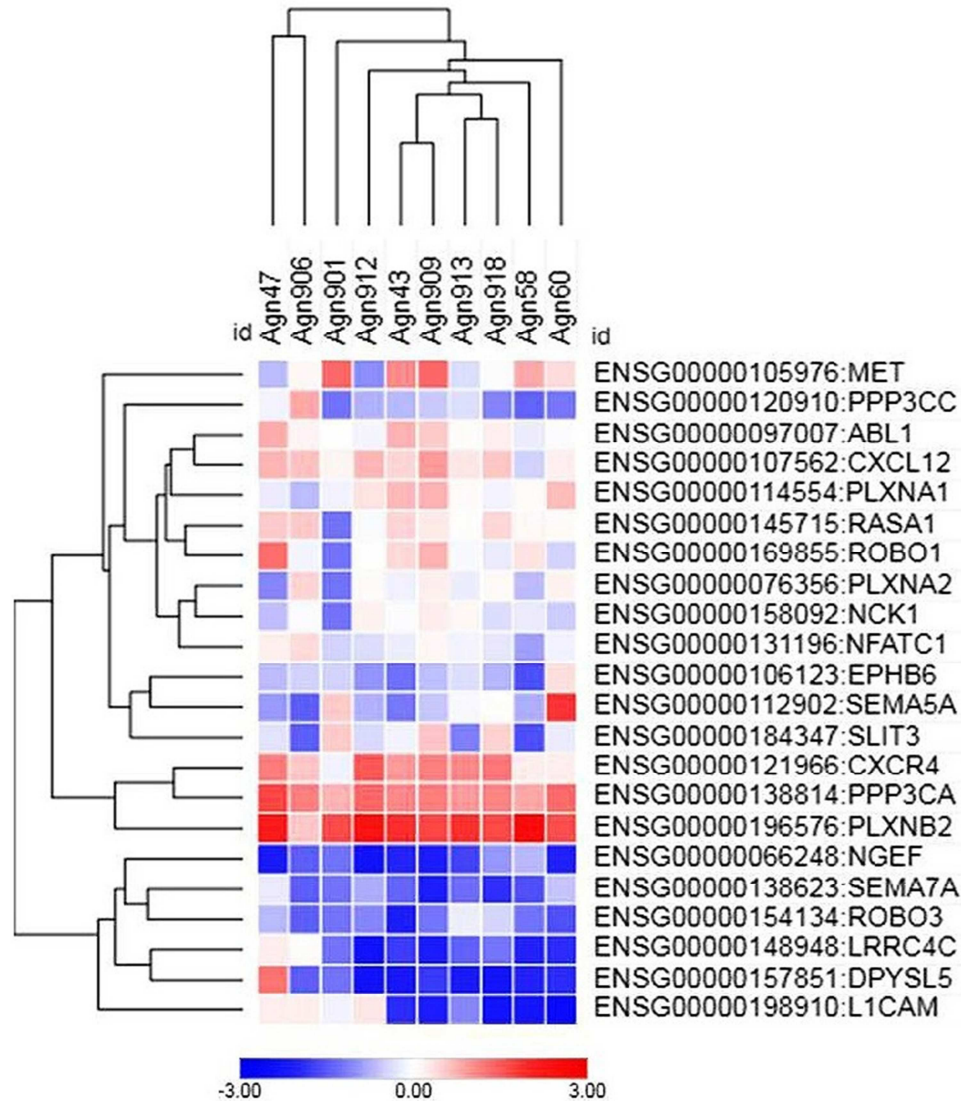


Figure 1. Heatmap representing a hierarchical clustering analysis of the expression profiles of axon guidance genes mutated in the CUP cohort.

By analyzing the transcriptomic profile of each tumor sample, z-score values were calculated, as  $\log_2(\text{cpm} + 1)$ . Mutated genes belonging to the Axon Guidance pathway (as annotated in KEGG pathway database) were then extracted and hierarchical clustering analysis was performed using Morpheus tool (<https://software.broadinstitute.org/morpheus/>), with the following setting parameters: Metric: Euclidean distance; Linkage method: Average.

mutations in axon guidance pathway genes were enriched in CUPs versus each of the 24 comparable tumor types and never under-represented; and this difference was highly statistically significant (with  $\text{FDR} < 0.1$  in all cases, but one with  $\text{FDR} = 0.103$ ). Moreover, over 70% of the samples in the CUP cohort, independent of histopathological features, actually carried mutations in genes of the axon guidance pathway, with one or often multiple hits, and up to 21 genetic changes in some cases (Table EV1; sequencing data deposited in the European Genome-phenome Archive, under the accession code EGAS00001004868).

In order to tackle the potential functional involvement of these mutated genes in the CUP phenotype, we complemented WES analysis with gene expression profiling. The analysis was performed on 10 out of 14 samples for which RNA was available, and we focused

our attention on axon guidance genes found to be mutated in CUPs (Fig 1). Based on Z-score determination in each sample, a small cluster of three genes resulted prominently expressed in all CUPs: *CXCR4*, *PPP3CA* and especially *PLXNB2*, suggesting that their mutated transcripts are likely to be expressed and potentially impact on CUP cell behavior.

A *PLXNB2* mutation found in our CUP cohort, and not reported previously in COSMIC, encoding p.G842C amino acid change, was particularly intriguing based on our structural *in silico* analysis (detailed below). Moreover, this change was found to be conserved in 14 independent metastatic lesions biopsied in the CUP patient AGN43, as well as in the matched patient-derived *ex-vivo* model, called agnospheres (Verginelli *et al.*, 2021). Two other *PLXNB2* mutations (encoding R531P and L1058S amino acid changes) were



		3° IPT domain			
<b>B2MUT</b>	803	F	VITRIQ <b>E</b> ET <b>G</b> PLGGGIRI <b>T</b> ILGSNLGVQAGDI-QRIS <b>V</b> AC <b>R</b> NC <b>S</b> FQPERY <b>S</b> VSTRIV <b>C</b>	860	
<b>B2</b>	803	F	VITRIQ <b>E</b> ET <b>G</b> PLGGGIRI <b>T</b> ILGSNLGVQAGDI-QRIS <b>V</b> AG <b>R</b> NC <b>S</b> FQPERY <b>S</b> VSTRIV <b>C</b>	860	
<b>B1</b>	1070	P	LIHSVE <b>E</b> LT <b>G</b> PVDGGTRV <b>T</b> IRGSNLGQHVDVLMVT <b>V</b> AG <b>V</b> PC <b>A</b> VD <b>A</b> Q <b>E</b> Y <b>E</b> VSSSLV <b>C</b>	1128	
<b>B3</b>	856	F	SIDAVE <b>E</b> LT <b>G</b> P <b>E</b> GG <b>L</b> AL <b>T</b> ILGSNLGRAFADVQYAVSVASRP <b>C</b> NPEPSLY <b>R</b> TSARIV <b>C</b>	916	
<b>A1</b>	864	P	KILKLS <b>E</b> ET <b>G</b> PRQGG <b>T</b> RL <b>T</b> ITGENLGLRFEDVRLGVRVGKVL <b>C</b> SPVSESE <b>Y</b> ISAEQIV <b>C</b>	922	
<b>A2</b>	858	F	QITEILTVS <b>G</b> P <b>E</b> GG <b>T</b> RV <b>T</b> IHG <b>V</b> NLGLDFSEIAHHVQ <b>V</b> AG <b>V</b> PC <b>T</b> PLPGE <b>Y</b> IIAEQIV <b>C</b>	916	
<b>A3</b>	810	P	RI <b>T</b> QI <b>H</b> EL <b>V</b> GP <b>K</b> EG <b>G</b> TRV <b>T</b> IVG <b>E</b> NLGLLSREVG--LR <b>V</b> AG <b>V</b> RC <b>N</b> SI <b>P</b> AE <b>Y</b> ISAERIV <b>C</b>	866	
<b>A4</b>	858	F	RI <b>T</b> EI <b>I</b> EV <b>T</b> GP <b>R</b> EG <b>G</b> TKV <b>T</b> IRG <b>E</b> NLGL <b>E</b> FRDIASHVK <b>V</b> AG <b>V</b> EC <b>S</b> PLVDG <b>Y</b> IPAEQIV <b>C</b>	916	
<b>D1</b>	891	P	EIHAI <b>E</b> EL <b>S</b> GPLDGG <b>T</b> LL <b>T</b> IRGRNLGRRLSDDVAHG <b>V</b> WIG <b>G</b> V <b>A</b> CEPLPDR <b>Y</b> TVSEEV <b>C</b>	949	
<b>Met</b>	657	F	VITSI <b>S</b> E <b>K</b> Y <b>G</b> PMAG <b>G</b> TL <b>L</b> LTG-NYL--NSGNSRHISIG <b>G</b> KT <b>C</b> TLK---SVSN <b>S</b> ILE <b>C</b>	709	

Figure 2. Multiple alignment of the primary sequences of IPT domains found in Plexin family members and in the homologous Met receptor.

In the first line is reported the G842C-mutated sequence of IPT3 domain of PlexinB2, found in CUP samples. The primary sequences of corresponding IPT3 domains in plexin family members, and in Met receptor tyrosine kinase, are aligned. Nucleotide positions at start/end of the nucleotide stretches are indicated. Highlighted in yellow are highly conserved amino acid residues. Cysteines engaged in disulfide bonds are underlined. The conserved glycine mutated in PlxnB2 is indicated in bold, as well as neighboring cysteines potentially engaged in covalent links.

found in independent CUP samples, from which it was not possible to derive *ex vivo* primary cancer cell models. We therefore decided to focus our study on the potential pathogenic role of this mutant axon guidance receptor in CUP cells.

G842C mutation of PlxnB2 replaces a conserved residue in the extracellular domain, with predicted impact on receptor conformation

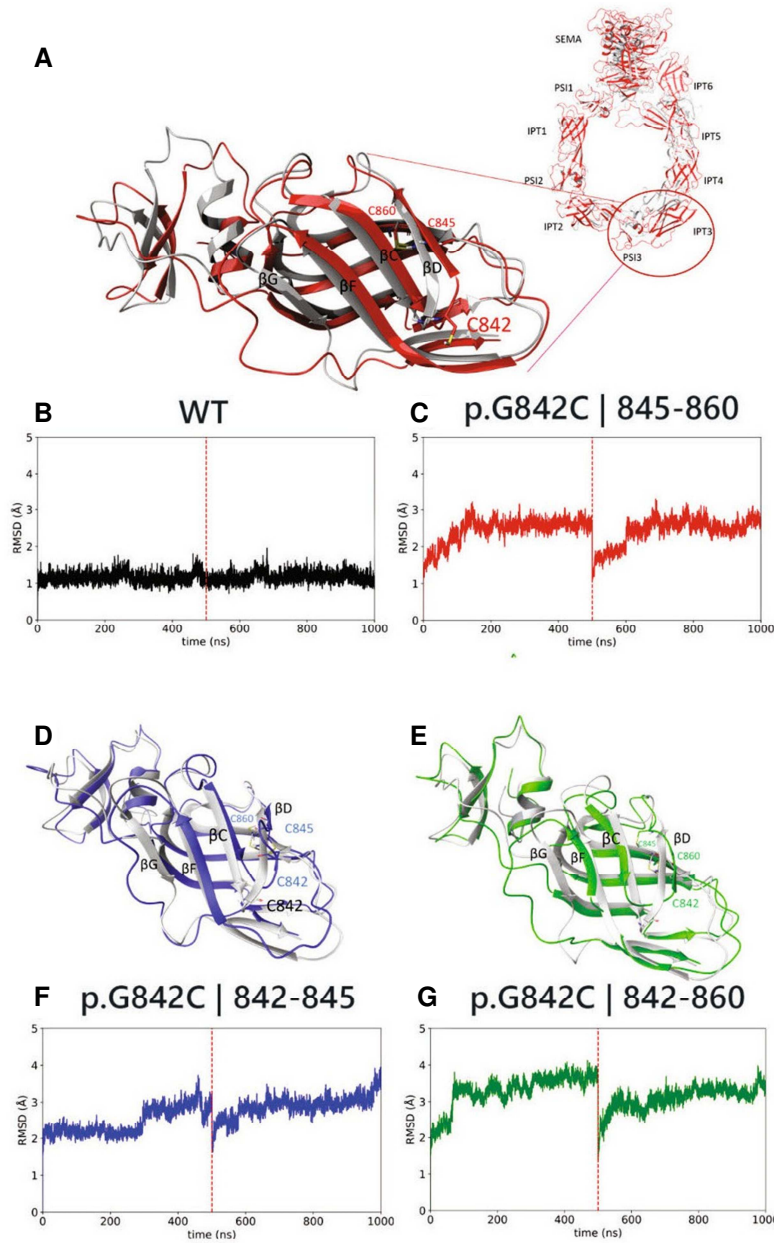
PlxnB2 belongs to the plexin family of semaphorin receptors, which comprises nine members in humans. The extracellular moiety of these single-pass transmembrane proteins includes a conserved sema domain, followed by series of cysteine-rich PSI motifs and IPT domains; the latter are characterized by an immunoglobulin-like fold and considered a protein–protein interaction interface. *In silico* analysis of PlxnB2 primary amino acid sequence revealed that Gly842 residue is located in the frame of the conserved third immunoglobulin-like IPT domain. This amino acid site is topologically conserved in six out of nine members of the human plexin family; moreover, it is located upstream of two cysteines (in +3 and +18 position), which are conserved in the corresponding IPT domain of all plexins, and known to be linked by a disulfide intramolecular bond (Fig 2). A similar structure is found in the homologous first IPT domain of Met and Ron tyrosine kinase receptors, which were previously found to be affected by activating mutations in human cancers (Ma *et al*, 2010; Navis *et al*, 2015).

In order to predict the potential damaging impact of G842C amino acid change in PlxnB2, the structure of this protein was retrieved from AlphaFold database (Jumper *et al*, 2021; Varadi *et al*, 2022) and three-dimensional model structures of human wild-type plexin and of G842C variant have been generated. Molecular dynamics (MD) simulations were then run to challenge the system. As shown in Fig 3A (focusing on the IPT3 domain), the superimposition of the wild-type protein so obtained (gray structure) with the selected variant (red structure) demonstrates that this mutation does

not alter significantly the global fold of PlxnB2 quaternary structure. At the same time, the results of two consecutive MD runs, of 500 ns each, clearly revealed that the mutation caused structural instability of the IPT3 protein domain (Fig 3B and C), as revealed by an increase in root-mean-square deviation of atomic position (RMSD) values with respect to the wild type, along both measured trajectories. Such tertiary structure variation occurs as the substitution of the small glycine residue in this position slackens the loop between  $\beta$ C and  $\beta$ D strands, entailing a greater flexibility on the entire IPT3 domain, and the eventual unfolding of  $\beta$ G strand in the G842C mutant (red structure) with respect to the wild-type protein (gray structure).

Notably, as discussed previously, the variant Cys842 residue is located at a binding distance with both Cys845 and Cys860 residues; therefore, an equilibrium among alternative intrachain disulfide bonds is conceivable. Accordingly, two alternative structures have been analyzed, challenging the formation of disulfide bridges between residues 842–845 (G842C | 842–845) or 842–860 (G842C | 842–860), and two MD simulations of 500 ns each were run for any structure determined. In Fig 3D and E are shown the superimpositions between the wild-type protein structure (gray) and the predicted structure of MD simulations for G842C | 842–845 (blue) or G842C | 842–860 (green), respectively. It is noteworthy that these potential changes in the native fold are not subverting the global structure of PlxnB2, but lead to further destabilization of the tertiary structure of the IPT3 domain, as demonstrated by increased RMSD values compared with both the wild-type and the mutant p.G842C structure shown above (Fig 3F and G). Again, the main culprit is the disruption of the  $\beta$ G strand; but here, due to alternative disulfide bond formation, this is also accompanied by repositioning and partial unfolding of the  $\beta$ D strand, thus enhancing the divergence in BETARMSD score with respect to the wild-type protein (Fig EV1).

In order to validate the relevance of our *in silico* predictions about the impact of G842C mutation, we analogously assessed the consequence of introducing the other PlxnB2 mutations found in



**Figure 3.** Structural molecular modeling of PlxnB2 wild-type and G842C-mutated form.

A, D, E Most representative conformations of IPT3 domains of PlxnB2 p.G842C (A, in red), PlxnB2 p.G842C 842–845 (D, in blue) and PlxnB2 p.G842C 842–860 (E, in green) compared with WT most representative conformation (in gray).

B, C, F, G RMSD plots of the concatenated replicas of the 4 mentioned IPT3 domains [PlxnB2 WT, p.G842C, p.G842C/842–845, and p.G842C/842–860, respectively].

CUP samples, R531P and P1058S, as well as mutations in the IPT3 domain of PlxnB2 reported in COSMIC and TCGA databases: R820H, L828F, R843Q, and Y852C (see [Materials and Methods](#) for data source). This analysis revealed that none of the other mutations had a comparable impact as G842C on the tridimensional conformation of the protein. Notably, simulations for few other mutations falling in the IPT3 domain of PlxnB2 (i.e., R820H, R843Q, and Y852) revealed modest but appreciable alterations of the RMSD plots (Fig EV2).

G842C is a putative gain-of-function mutation leading to PlxnB2 activation in ligand-independent manner

Plexins are well-known receptors for semaphorins, and the prime ligand for PlxnB2 is Sema4C (Deng *et al*, 2007). Notably, semaphorin-induced plexin signaling is often associated with the induction of the so-called “collapsed” phenotype, characterized by retraction of cellular processes and cell rounding; this is a typical *in vitro* response, which is considered a surrogate indicator of plexin



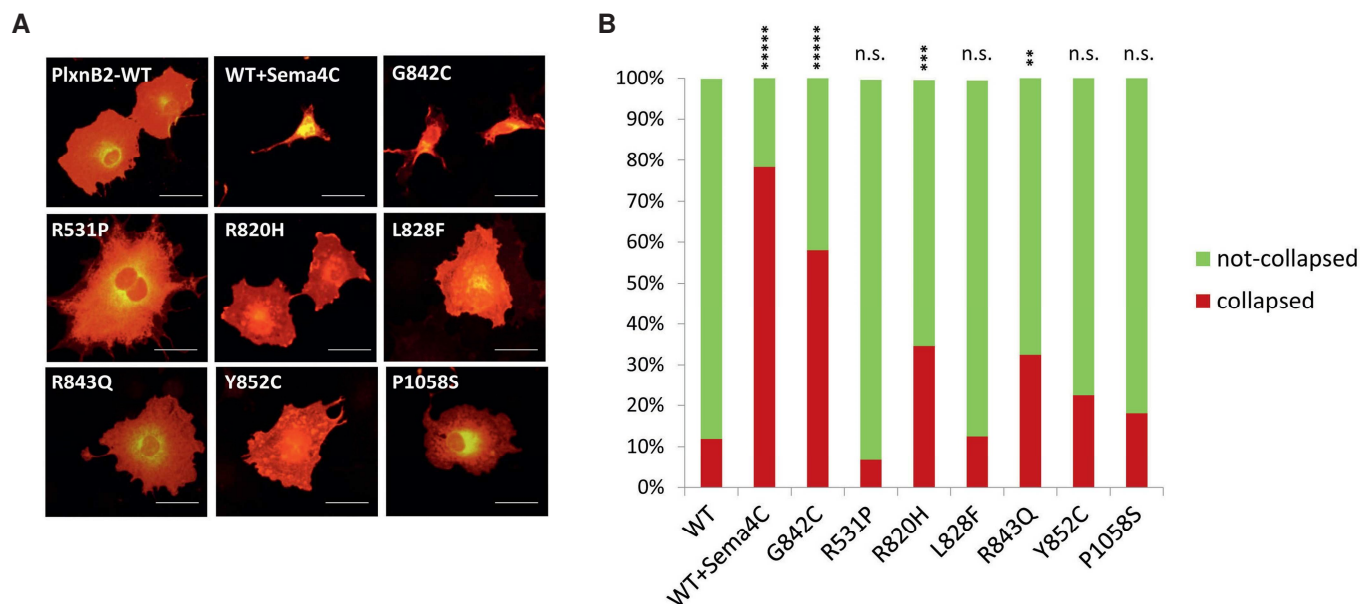


Figure 4. G842C mutation leads to spontaneous activation of PlxnB2 signaling.

A Representative images of immunofluorescence analysis (with anti-PlxnB2 antibodies) of the phenotype of COS7 cells transfected with wild-type or mutated PlxnB2, compared with the collapsed phenotype induced by the cognate ligand Sema4C (1  $\mu$ g/ml) in cells expressing the WT receptor. Scale bar: 50  $\mu$ m. A wide series of representative microscopic images is further shown in Appendix Fig S2.

B The stacked bar graph shows the fraction of collapsed PlxnB2-expressing cells, identified as having a surface minor of 2000  $\mu$ m<sup>2</sup> (measured by ImageJ software), out of 50–100 cells counted per each condition, from different fields and  $n > 3$  experiments). The statistical analysis was performed comparing WT untreated cells with all other groups, by Fisher's exact test, which indicated a \*\*\*\* $P$ -value  $< 0.00001$  for Sema4C-treated and for G842C-PlxnB2-expressing cells (confirmed by ANOVA analysis on individual cell values, shown in Appendix Fig S2). By Fisher's test, the only other mutants significantly different from WT were R820H ( $P < 0.001$ ) and R843Q ( $P < 0.01$ ), although ANOVA analysis on individual cell values shown in Appendix Fig S2 did not confirm such significance for other PlxnB2 variants.

Source data are available online for this figure.

activation. Intriguingly, COS7 cells expressing G842C-mutated PlxnB2 displayed a prevalent collapsed phenotype in the absence of the ligand, comparable to that induced by Sema4C in cells bearing the wild-type receptor, which are otherwise mainly spread in basal conditions (Fig 4A and B and Appendix Fig S2). Notably, cells overexpressing the other PlxnB2-IPT3 mutants described above did not show significant changes in the basal phenotype or displayed a modest increase in the number of collapsed cells, significantly inferior to that elicited by G842C mutation discovered in CUPs (Fig 4A and B and Appendix Fig S2). We concluded that G842C-PlxnB2 acts as a putative gain-of-function mutation, leading to ligand-independent receptor activation in overexpressing cells. Noteworthy, although the IPT3 domain of plexins is presumptively not implicated in ligand binding, we also verified whether G842C-PlxnB2 mutation may affect the interaction with Sema4C. Indeed, in a classical *in situ* binding assay with alkaline-conjugated Sema4C, the interaction with the mutated receptor expressed in COS cells appeared unchanged, compared with the wild-type counterpart (Appendix Fig S3).

Knocking down G842C-mutated, but not wild-type PlxnB2, hampers CUP cell viability, clonogenic capacity, and tumorigenic growth in mice

In order to assess the functional relevance of endogenous G842C-mutated PlxnB2 protein, we knocked down its expression in

agnospheres AS43, derived from CUP patient AGN43 (validation data in Fig EV3), and evaluated cell proliferation and stem cell frequency. As a comparison, we performed similar experiments in agnospheres AS901, AS906, or AS67 (derived from patients AGN901, AGN906, and AGN67, respectively), which express wild-type PlxnB2. As shown in Fig 5A, PlxnB2 knockdown had no significant impact on the growth of agnospheres carrying wild-type PlxnB2 (compared with controls expressing scrambled shRNAs); in sharp contrast, the depletion of G842C-mutated PlxnB2 in AS43 resulted in a striking inhibition of cell viability and growth versus the respective controls. Similar results were obtained upon PlxnB2 depletion by means of another independent shRNA sequence (Fig EV3).

To assess whether G842C-PlxnB2 could have an impact on cancer cell self-renewal, we performed limiting dilution assays (LDA). By ELDA software analysis, the stem cell frequency (featuring the fraction of clonogenic cells) was estimated in each experimental condition. Consistent with previous observations (Verginelli *et al*, 2021), LDA demonstrated a high clonogenic capacity of AS43 control cells (Fig 5B). Interestingly, the stem cell frequency was strongly reduced in PlxnB2-depleted AS43. By contrast, no significant differences were observed in AS901, AS906, or AS67, upon silencing the wild-type PlxnB2 receptor. Altogether, these data suggest that AS43 cancer cells are specifically dependent on G842C-mutated PlxnB2 for stem properties and proliferation.

Stemming from the results above, to better investigate the role of mutated-PlxnB2 in tumor progression, we generated agnosphere-

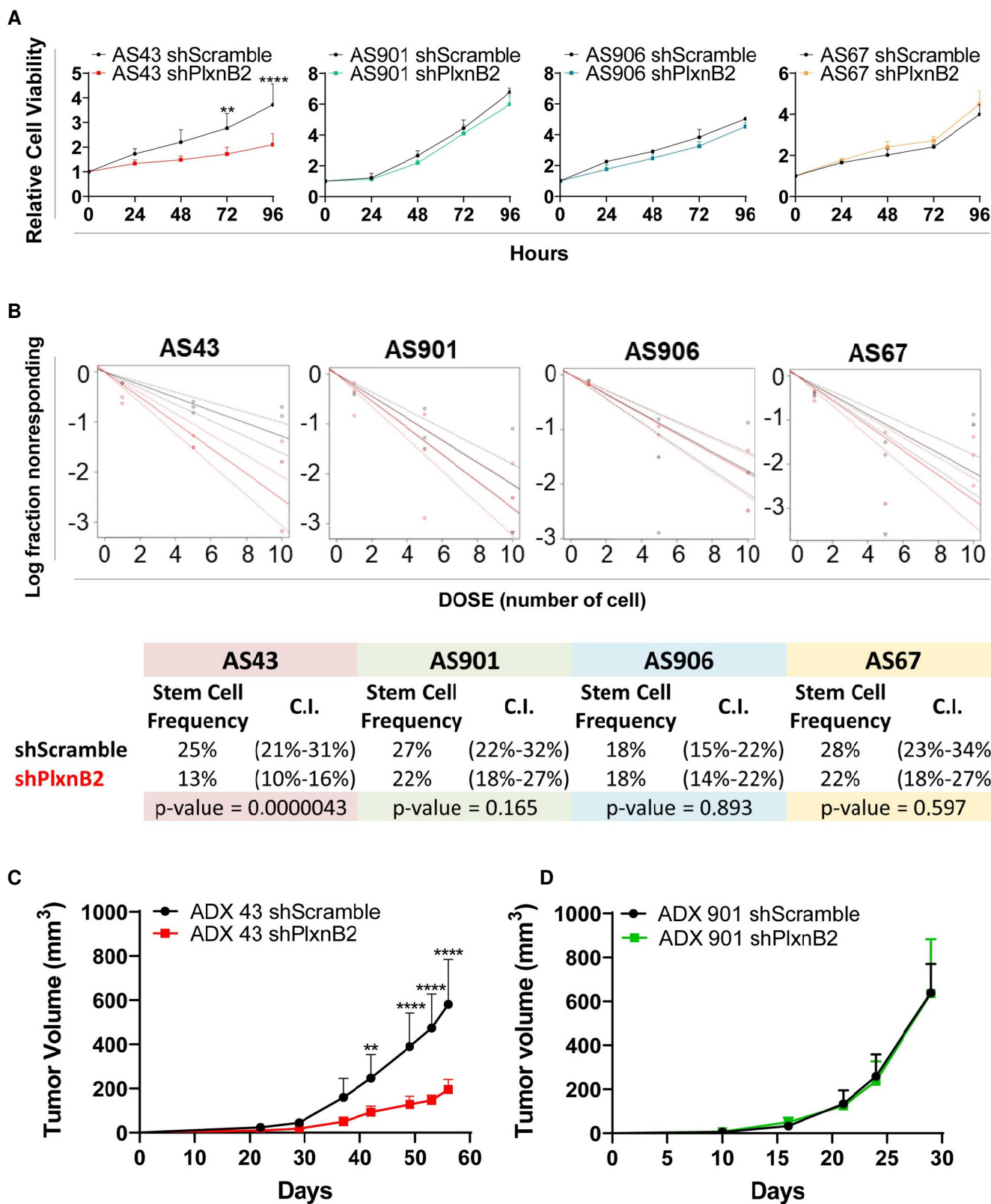


Figure 5.

Figure 5. Knocking down G842C-mutated, but not wild-type PlxnB2, hampers CUP cell viability, clonogenic capacity, and tumorigenic growth in mice.

- A The graphs show the analysis of cellular viability in shPlxnB2-transduced agnospheres, compared with their respective scramble-transfected controls, after 24, 48, 72, and 96 h of growth in culture. Values are mean  $\pm$  SD of  $n \geq 3$  independent experiments (6 technical replicates for each). The statistical significance was assessed by two-way ANOVA with Sidak's correction for multiple comparisons: \*\* $P = 0.0053$ ; \*\*\*\* $P < 0.0001$ .
- B Limiting dilution assays (LDA, see details in [Materials and Methods](#)) were used to assess the clonogenic capacity of dissociated CUP cells derived from the same shPlxnB2-transduced agnospheres analyzed above, compared with their respective controls. Analyses generated by the ELDA software are shown below, reporting the estimated stem cell frequency (percentage of clonogenic cells) with confidence intervals (C.I.) and statistical analysis by chi-square test.
- C, D (C) After transplantation in immunodeficient mice of CUP cells described above, the volume of ADX43- and (D) ADX901-shPlxnB2 tumors was assessed, in comparison with the respective controls. Plotted values indicate the mean  $\pm$  SD ( $n = 8$  each group). The statistical significance was assessed by two-way ANOVA with Sidak's correction for multiple comparisons: \*\* $P = 0.0060$ ; \*\*\*\* $P < 0.0001$ .

Source data are available online for this figure.

derived xenograft (ADX) preclinical models, as described previously (Verginelli *et al*, 2021); thus, we injected subcutaneously in NOD-SCID mice either control or PlxnB2-depleted AS43 or AS901 CUP models and monitored tumor growth by periodical calibration. Our results indicated that the expression of G842C-mutated PlxnB2 is specifically required for the growth of AS43-derived tumors. In fact, whereas we observed a marked reduction in tumor volume in PlxnB2-silenced ADX43 mice (Fig 5C), there was no significant difference between silenced and control tumors in ADX901 that carry wild-type PlxnB2 (Fig 5D), further suggesting that G842C-mutated (but not wt) PlxnB2 is promoting CUP-derived ADX growth *in vivo*.

#### G842C-mutated PlxnB2 enhances CUP cell proliferation and tumor growth *in vivo*

Next, to determine whether the G842C mutation can actually drive tumor cell proliferation, we overexpressed wild-type or mutated PlxnB2 in AS901, AS906, and AS67, which do not basally carry the mutation (validation data in Fig EV3). Interestingly, we found enhanced proliferation of the cells overexpressing mutated PlxnB2 compared with both wt-PlxnB2 and control mock-transduced agnospheres (Fig 6A and Fig EV4).

To further explore the potential of G842C PlxnB2 mutation in a preclinical model, we inoculated NOD/SCID mice with AS901 agnospheres overexpressing mutated or wild-type PlxnB2 (or mock controls). We observed enhanced growth in tumors overexpressing G842C-mutated PlxnB2 compared with both PlxnB2 wt and controls (Fig 6B). Altogether, these data support the conclusion that the mutant G842C-PlxnB2 receptor drives a mechanism promoting the tumorigenic potential of CUP cells.

#### G842C-mutated PlxnB2 enhances cancer cell migration and invasion

In a previous study, we demonstrated that higher PlxnB2 signaling was associated with increased MCF7 cancer cell migration (Gurrapu *et al*, 2018). These luminal-type breast cancer cells are rather indolent and noninvasive. Thus, in order to assess a potential impact of the G842C mutation in plexin-dependent invasiveness, we initially overexpressed wild-type or mutated-PlxnB2 in MCF-7 cells (Fig 7A). PlxnB2 overexpression, either wild-type or mutated, did not affect the growth rate of MCF7 cells propagated in culture in the presence of serum (unpublished observation, by Serena Brundu). Consistent with previous findings, PlxnB2-WT overexpression promoted basal

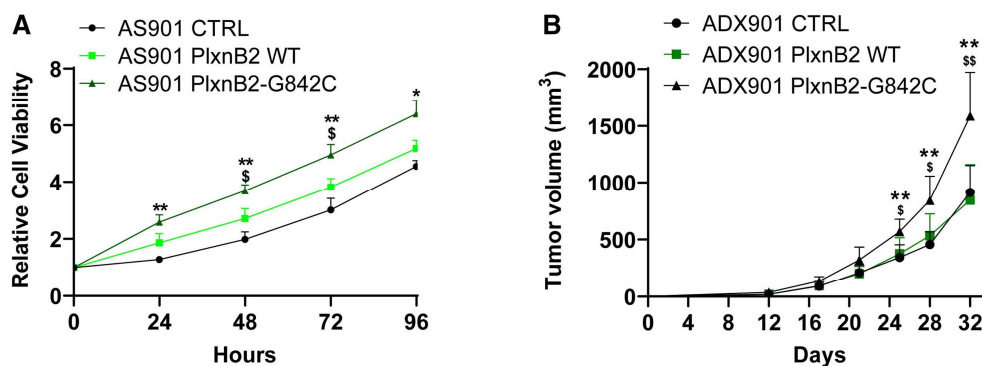


Figure 6. Overexpression of G842C-mutated PlxnB2 in wild-type CUP cells enhanced AS viability and tumor growth in mice.

- A Time course analysis of the cellular viability of AS901, stably overexpressing either wild-type or G842C-mutated PlxnB2, or controls transduced with an empty vector. Plotted values are the mean  $\pm$  SD of  $n = 3$  independent experiments (six technical replicates for each). The statistical significance was assessed by two-way ANOVA with Tukey's correction for multiple comparisons. PlxnB2-G842C vs. controls (CTRL): at 24 h \*\* $P = 0.0097$ , 48 h \*\* $P = 0.0026$ , 72 h \*\* $P = 0.0079$ , 96 h \* $P = 0.0226$ . PlxnB2-G842C vs. PlxnB2-WT: at 48 h  $^{\$}P = 0.0418$ , 72 h  $^{\$}P = 0.0318$ .
- B Tumor growth curves of ADX 901 overexpressing either PlxnB2-G842C or PlxnB2-WT, or mock control. Plotted values are the mean  $\pm$  SD ( $n = 8$  each group). The statistical significance was assessed by two-way ANOVA with Tukey's correction for multiple comparisons. PlxnB2-G842C vs. control (CTRL): at 25 days \*\* $P = 0.0029$ , 28 days \*\* $P = 0.0028$ , 32 days \*\* $P = 0.0033$ . PlxnB2-G842C vs. PlxnB2-WT: at 25 days  $^{\$}P = 0.0221$ , 28 days  $^{\$}P = 0.0234$ , 32 days  $^{\$}P = 0.0024$ .

Source data are available online for this figure.

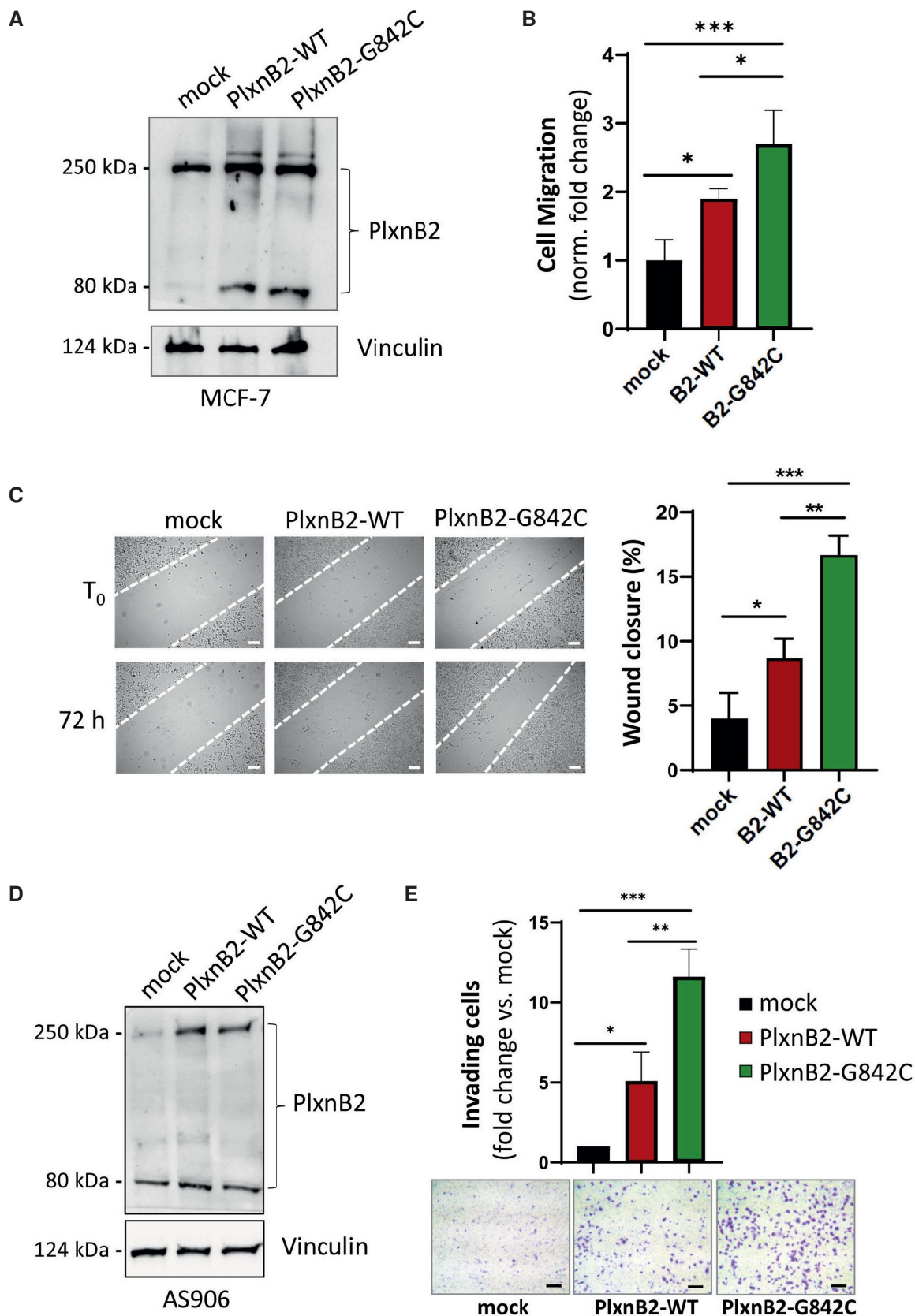


Figure 7.



**Figure 7.** Overexpression of G842C-PlxnB2 promotes cancer cell migration/invasion.

- A Western blotting showing basal endogenous or overexpressed PlxnB2, either WT or G842C-mutated isoforms, in MCF7 cells. In this wide view of the hybridized membrane, it is appreciable the presence of full-length molecules at 250 kDa, as well as the processed 80 kDa fragments (reported in Artigiani *et al.*, 2003). Vinculin staining provided protein loading controls.
- B The graph shows the mean  $\pm$  SD of relative increase in 24-h spontaneous migration across Transwell inserts of MCF7 cells (analyzed above) overexpressing either wild-type or G842C-mutated PlxnB2, versus mock-transfected controls ( $n = 4$  biological replicates). The statistical significance was verified by one-way ANOVA with Bonferroni correction for multiple comparisons: PlxnB2-WT vs. mock  $*P = 0.0145$ ; PlxnB2-G842C vs. PlxnB2-WT  $*P = 0.0277$ ; PlxnB2-G842C vs. mock  $***P = 0.0002$ . Representative images of the cells migrated across the semipermeable membrane are shown in Appendix Fig S4A.
- C Wound healing scratch assays on monolayers of MCF7 cells engineered as described above. The graph shows the mean  $\pm$  SD of values obtained in  $N = 3$  independent experiments (with at least three technical replicates for each). The statistical significance was verified by one-way ANOVA with Tukey's correction for multiple comparisons:  $*P = 0.0350$ ;  $**P = 0.0029$ ;  $***P = 0.0002$ . Representative phase contrast images of the wounded monolayers are further shown in larger size in Appendix Fig S4B.
- D Western blotting analysis of endogenous and overexpressed PlxnB2 (either WT or G842C-mutated isoforms) in CUP cells AS906 (both full-length molecules at 250 kDa and processed 80 kDa fragments). Vinculin staining provided protein loading controls.
- E The invasive capacity of CUP cells AS906 (shown in previous panel), overexpressing WT or G842-mutated PlxnB2 (or mock control), was assessed across Matrigel-coated Transwell inserts. The graph shows the mean  $\pm$  SD of values obtained in  $N = 3$  independent experiments (including technical duplicates, in two cases). Below the graph, are shown representative fields of the semipermeable membranes (additional larger size images can be found in Appendix Fig S4C); scale bar: 100  $\mu\text{m}$ . The statistical significance was verified by one-way ANOVA with Bonferroni correction for multiple comparisons:  $*P = 0.0406$ ;  $**P = 0.0045$ ;  $***P = 0.0003$ .

Source data are available online for this figure.

cancer cell migration to some extent; however, here we found that this effect was far more pronounced in the presence of G842C-mutated receptor (Fig 7B and Appendix Fig S4A). We furthermore analyzed MCF-7 cell migration by scratch assays, in which the gap of a “wounded” monolayer may be filled by migrating cells. Again, wound healing by cells transduced with wild-type PlxnB2 was basally increased compared with controls; however, the expression of G842C-PlxnB2 elicited a strikingly greater wound closure, further suggesting that the mutated plexin was associated with a greater functional activity (Fig 7C and Appendix Fig S4B).

The importance of G842-mutated PlxnB2 in cancer invasion was further assessed in the framework of CUP cells carrying an endogenous wild-type gene. Thus, AS906 transduced to overexpress either wild-type or mutated PlxnB2 (or control cells; Fig 7D) were seeded in the upper chamber of a Transwell insert, on the top of a Matrigel layer, recapitulating the extracellular matrix (ECM). Notably, in the presence of G842-mutated PlexinB2, CUP cells were consistently more efficient than controls in invading the ECM barrier, indicating that this mutation mediates a gain-of-function highly relevant for cancer progression (Fig 7E and Appendix Fig S4C).

#### G842C-mutated PlxnB2 is required and sufficient to promote EGFR phosphorylation and EGFR-dependent invasiveness in CUP cells

We aimed at identifying signaling mechanisms potentially responsible for the promotion of CUP cell stemness, proliferation, and invasiveness by G842C-mutated PlxnB2. It was previously shown that ligand-dependent PlexinB2 activity elicits ErbB2 tyrosine phosphorylation and oncogenic signaling in breast cancer cells (Gurrupu *et al.*, 2018). However, we did not observe a significant change in phospho-ErbB2 levels in association with G842C-mutated PlxnB2 expressed in CUP cells (unpublished observation, by Virginia Napolitano). Intriguingly, in a previous study, we reported that the homologous EGFR tyrosine kinase is basally phosphorylated at low levels in CUP cells carrying a wild-type PlxnB2 (AS901, AS906), while it is highly phosphorylated in AS43 cells, which carry the G842C-mutated receptor (Verginelli *et al.*, 2021). EGFR is a major oncogenic promoter in human cancers, especially upon gene

amplification and overexpression (Sigismund *et al.*, 2018); however, AS43 do not carry *EGFR* mutations or elevated expression, so we posited that the presence of G842C-PlxnB2 might promote its noncanonical activation.

First, we tested the impact of silencing the mutated-PlexinB2 in AS43 CUP cells and found that this resulted in strikingly reduced EGFR tyrosine phosphorylation (Fig 8A), confirming that this altered axon guidance cue controls a major oncogenic pathway in CUP cells. We then analyzed phospho-EGFR levels in basally nonmutated CUP cells (AS906 and AS901), upon transduction of either the wild-type or the G842C-mutated PlxnB2 isoform expressed in AS43. Data in Fig 8B show that the mutated plexin elicited a significantly stronger EGFR phosphorylation, with respect to controls. Thus, our data indicate that G842C-PlxnB2 expression is required and sufficient to sustain EGFR phosphorylation in CUP-derived agnospheres.

Although we were unable to co-immunoprecipitate endogenous EGFR and G842C-PlxnB2 in CUP cells, the two receptors strongly and specifically associated in a complex upon transient transfection in HEK293T cells (Fig EV5A); in this setting, however, we could not achieve conclusive evidence about a quantitative difference between EGFR association with mutated vs. wild-type PlxnB2. Thus, our data suggest that the two receptors can interact in a dynamic complex on the cell surface, which could enable EGFR trans-regulation by mutation-driven conformational changes in PlxnB2.

In order to assess whether the increased basal activation of EGFR in CUP cells carrying mutated PlxnB2 might be accounted to explain their proliferative self-sufficiency, we functionally inhibited EGFR by the specific monoclonal antibody cetuximab, commonly used as anticancer drug. Notably, EGFR blockade in AS43 was associated with abrogated ERK and S6 phosphorylation (Fig 8C), indicating that these intracellular effectors are dependent on EGFR signaling in cells carrying mutated PlexinB2. By contrast, the same pathways were unaffected by cetuximab treatment in CUP cells expressing wild-type PlxnB2 (Fig 8C).

In line with these findings, cetuximab treatment resulted in a substantial inhibition of AS43 cell viability and clonogenic capacity (Fig 9A–C), indicating that these cells are functionally dependent on EGFR signaling. By contrast, AS901, AS906, and AS67, which carry wild-type PlxnB2, were minimally affected or unresponsive to EGFR



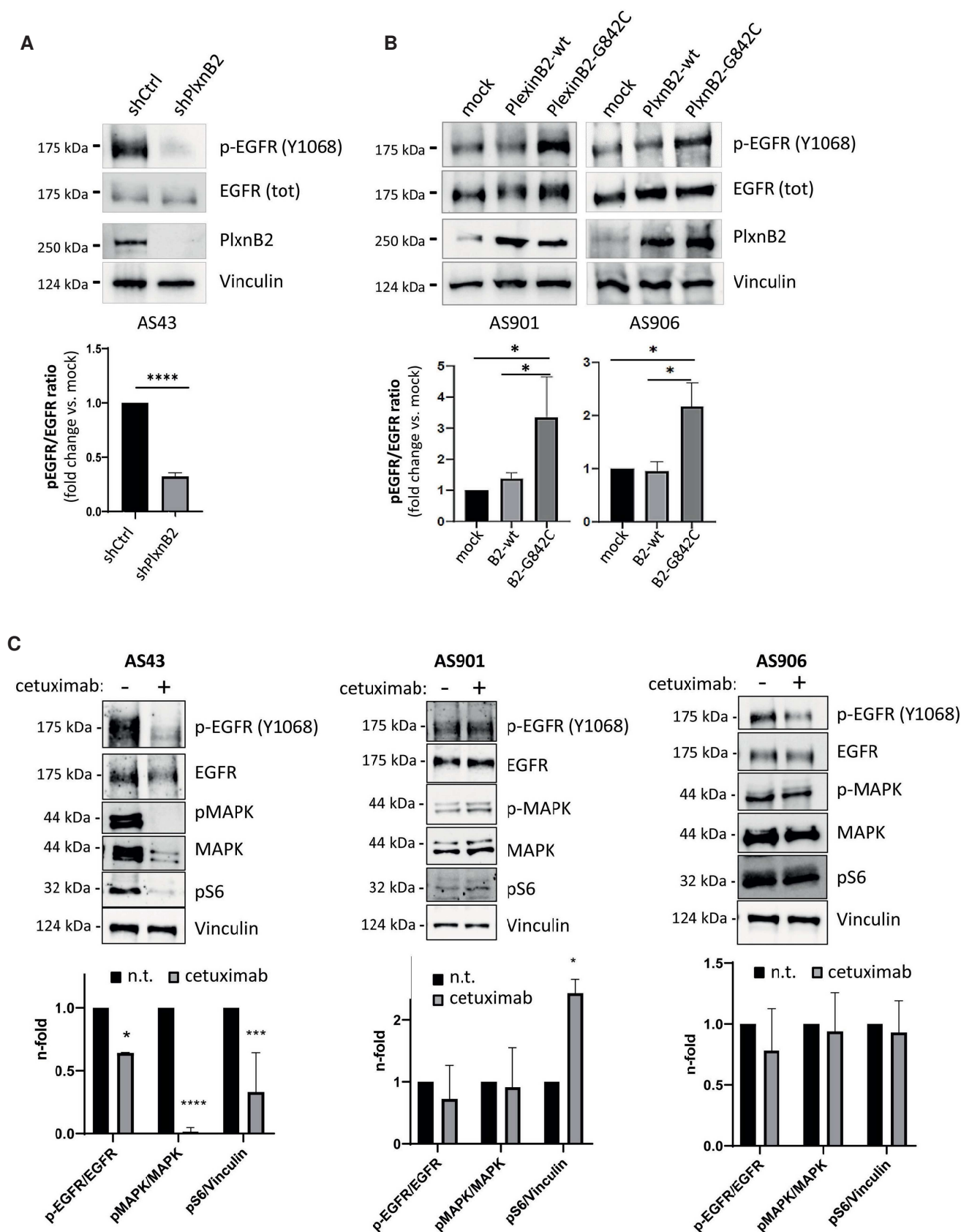


Figure 8.

**Figure 8.** Expression of G842C-PlxnB2 controls EGFR phosphorylation and intracellular signaling.

- A Phospho-EGFR analysis by western blotting in control and PlxnB2-silenced AS43 cells. Images show representative results of  $n = 3$  independent experiments, while the graph at the bottom shows the mean  $\pm$  SD of normalized band quantification. The statistical significance across replicates was verified by unpaired t-test: \*\*\*\* $P < 0.0001$ .
- B Phospho-EGFR analysis by western blotting in AS901 and AS906 cells either mock, or overexpressing WT or G842C-mutated PlxnB2. Images show representative results of  $n = 3$  independent experiments, while the graphs at the bottom show the mean  $\pm$  SD of normalized band quantification. The statistical significance across independent replicates was verified by one-way ANOVA followed by Tukey test: AS901 PlxnB2-G842C vs. PlxnB2-WT \* $P = 0.0436$ , PlxnB2-G842C vs. mock \* $P = 0.0211$ ; AS906 PlxnB2-G842C vs. PlxnB2-WT \* $P = 0.0436$ , PlxnB2-G842C vs. mock \* $P = 0.0481$ .
- C Western blotting analysis of pERK and pS6 intracellular signal transducers in the indicated agnospheres, after 6-h treatment with the EGFR inhibitor cetuximab (50  $\mu\text{g/ml}$ ). Images show representative results of  $n = 3$  independent experiments, while the graphs at the bottom show the mean  $\pm$  SD of normalized band quantification. The statistical significance across independent replicates was verified by Sidak's multiple comparison test between cetuximab-treated and untreated conditions, per each group: AS43 \* $P = 0.0153$ , \*\*\* $P = 0.0001$ , \*\*\*\* $P < 0.0001$ ; AS901 \* $P = 0.0201$ .

Source data are available online for this figure.

blockade (Fig 9A–C). Thus, the increased EGFR activity driven by genetically altered PlxnB2 controls intracellular signaling and proliferative self-sufficiency in CUP cells.

We finally asked whether EGFR kinase activity might be responsible for the increased invasiveness of CUP cells driven by G842C-mutated PlxnB2 overexpression. Notably, cetuximab treatment did not significantly interfere with AS906 cell basal invasion, indicating that this is not dependent on EGFR activity. By contrast, the elevated CUP cells invasiveness elicited by PlexinB2 G842C-mutated isoform was almost completely abrogated by the EGFR inhibitor (Fig 9D), indicating that the axon guidance receptor PlexinB2 is functionally placed upstream EGFR kinase in the regulation of CUP cells. Similar results were obtained upon EGFR blockade with the selective kinase inhibitor erlotinib (Fig EV5B).

## Discussion

Surveying the genomic mutational landscape of human cancers has revealed a number of driver genetic changes and associated dependencies on activated signaling cascades, enabling the development of specific targeted therapies. This is epitomized by the application of ErbB2-targeted drugs in a subset of breast cancers, of BRAF inhibitors in melanomas, of drugs targeting liabilities due to mutated BRCA in gynecological tumors, to cite some key examples. Notably, the therapeutic targeting of certain relevant mutations is still challenging, such as for K-RAS in colon and pancreatic cancers, but

thanks to progresses in research the goal is at sight. Yet, metastasis, the main culprit for cancer-associated lethality, could not be associated so far with specific genetic changes or driver pathways, disabling the chance to design targeted therapeutic approaches, and leaving scanty options for patient treatment at this stage. Paradigmatic examples of this scenario are cancers that appear as primarily metastatic at diagnosis; as such, they also miss the classical indications for tumor type-specific chemotherapy, which leads to the undertaking of agnostic therapeutic approaches with poor efficacy. Actually, unlike most human tumor types, the genomic mutational landscape of cancers of unknown primary (CUP) has been poorly investigated so far, mainly focusing on the hunt for genetic changes in known oncogenes and tumor suppressors.

In the present study, we performed a genome-wide unbiased analysis by NGS of a cohort of CUP samples. A limiting factor to this approach has previously been the lack of a coherent protocol to ensure the recruitment of definite CUP cases, accurately discriminated from early metastatic cancers for which the primary site had not been identified due to incomplete clinicopathological characterization. Thus, the CUP cohort analyzed here was unambiguously selected adopting current ESMO guidelines (Fizazi *et al*, 2015; Piscane *et al*, 2022). WES analysis of CUP samples did not reveal recurrent genetic changes in known oncogenes or tumor suppressors, apart from frequent hits in TP53, which are in the range of those observed in many other human tumors and can hardly be considered specific. A number of conclusions could be drawn from these findings; obviously, the lack of positive identification of specific

**Figure 9.** Viability and invasive capacity of CUP cells expressing G842C-mutated, but not wild-type, PlxnB2 is dependent on EGFR.

- A The cellular viability of the indicated agnospheres was assessed after 6 days in the presence or absence of the EGFR inhibitor cetuximab (50  $\mu\text{g/ml}$ ). In the plot, each box represents values from 25<sup>th</sup> to 75<sup>th</sup> percentiles, the central band indicates the mean, and the whiskers expand from maximum to minimum values of individual cetuximab-treated experimental points ( $n \geq 9$ ), normalized to the average value of untreated cells within each group. The statistical significance across three independent experiments, each comprising  $\geq 3$  technical replicates, was verified by unpaired t-test multiple comparisons of cetuximab-treated and untreated conditions, per each group: \*\*\* $P = 0.0002$ ; \*\*\*\* $P < 0.0001$ .
- B The viability of AS43, in presence (or absence) of cetuximab (50  $\mu\text{g/ml}$ ), was assessed over the indicated time course, compared with values at  $T_0$ . The graph shows the mean  $\pm$  SD of two independent experiments (with quadruplicate technical replicates each). The statistical significance was verified through multiple comparisons by two-way ANOVA with Bonferroni correction of cetuximab-treated and untreated samples, per each time point: \*\*\*\* $P < 0.0001$ .
- C Limiting dilution assays (LDA) were used to assess the clonogenic capacity of dissociated CUP cells derived from the indicated agnospheres in the presence or absence of cetuximab (50  $\mu\text{g/ml}$ ). Analyses generated by the ELDA software are shown below, reporting the estimated stem cell frequency (percentage of clonogenic cells) with confidence intervals (C.I.) and statistical analysis by chi-square test.
- D The invasive capacity of AS906 cells (overexpressing WT or G842C-mutated PlxnB2, or mock transfected) was assessed as in Fig 7E, in the presence or absence of cetuximab 50  $\mu\text{g/ml}$ . Images show representative results (scale bar: 100  $\mu\text{m}$ ), while the graph on the right shows the mean  $\pm$  SD of  $n = 3$  independent experiments. The statistical significance across replicates was verified by one-way ANOVA with Bonferroni correction: \*\* $P = 0.0010$ ; \*\*\* $P = 0.0005$ ; \*\*\*\* $P < 0.0001$ .

Source data are available online for this figure.

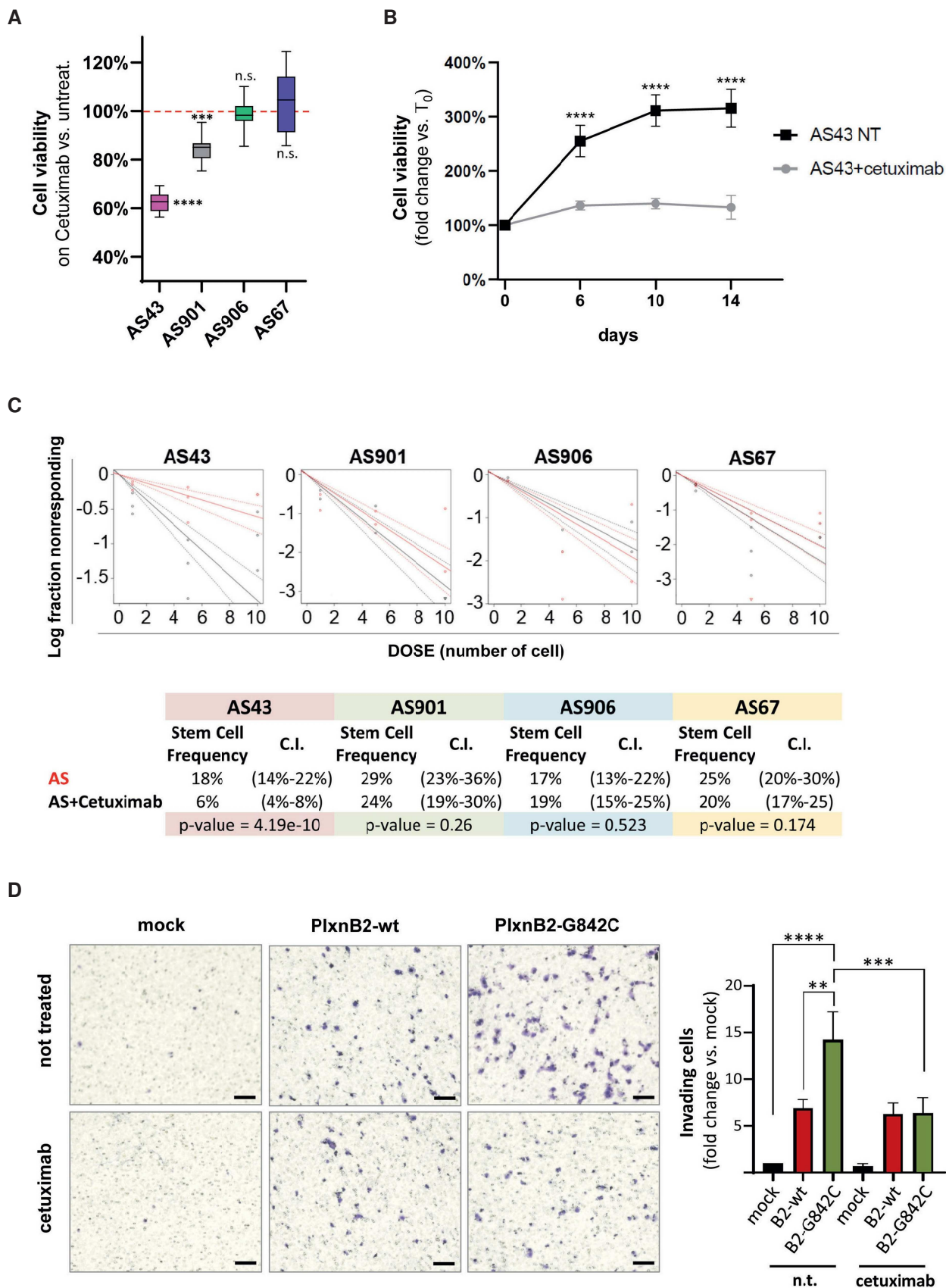


Figure 9.

metastasis-driver mutations may still be accounted by our inability to unveil their presence. However, we posited here that the CUP phenotype may also be sustained by the occurrence of diverse genetic (and possibly epigenetic) changes in some defined pathway controlling cell behavior. In order to tackle this hypothesis, we applied a validated algorithm calculating the frequency of mutations in genes that were associated in functional pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG). On these bases, the genomic mutational profile of CUPs was compared with those of 33 other tumor types (defined in TCGA database), revealing a significant enrichment of changes in specific pathways. In particular, mutations in axon guidance genes were found in 10 out of 14 CUP samples, independently from clinical or histopathological features of the tumors, and were more frequently occurring than in most other tumor types.

The axon guidance KEGG pathway comprises 125 genes, controlling cell migration and axonal navigation in development, but also associated with pathological dysregulation of cell motility in the adult stage. Notably, many of these genes have been implicated in tumor metastasis, including Met tyrosine kinase, semaphorins, and their receptors, the plexins. We decided to focus our attention on the mutated axon guidance gene most expressed in CUP samples, *PLXNB2*. Although previous studies have implicated this semaphorin receptor in the regulation of cancer cell proliferation, invasiveness, and metastatic spreading (Le *et al*, 2015; Yu *et al*, 2017; Gurrappu *et al*, 2018; Huang *et al*, 2021), none had associated its mutations with a causal role. In order to establish a proof of principle about the relevance of axon guidance genes in CUPs, we investigated G842C-PlxnB2 variant, which was suspected to be damaging based on sequence conservation and structural modeling results. Moreover, this mutation affected the conserved fold of an IPT domain, a moiety also found in Met and Ron oncogenic receptors and previously found to be affected by activating mutations in human cancers (Ma *et al*, 2010; Navis *et al*, 2015). Notably, the large intracellular portion of the plexins does not contain a kinase domain or other classical signaling domains, but was consistently reported to regulate the activity of monomeric GTPases, especially R-Ras, Rap-1, and RhoA, controlling cell migration, and axonal extension in neural development. In addition, plexins have been shown to couple with transmembrane tyrosine kinases of the receptor type such as ErbB2 and Met, thereby triggering alternative noncanonical signaling cascades, especially in cancer cells (Cagnoni & Tamagnone, 2014).

Our experimental evidence indicated that G842C-mutated PlxnB2 was competent for signaling, even in absence of semaphorin stimulation. Interestingly, other studies have recently reported signaling mechanisms mediated by plexins (including PlxnB2), which are not necessarily activated by semaphorins (Mehta *et al*, 2020; Jiang *et al*, 2021; Yang *et al*, 2022). Moreover, while knocking down PlxnB2 expression in CUP cells bearing a wild-type receptor had no apparent functional impact, the metastatic cells carrying G842C mutation proved to be dependent on this variant PlxnB2 to sustain self-renewal and proliferation in culture, as well as tumorigenesis in mice. Altogether, these data indicated that G842C-PlxnB2 can be considered a gain-of-function mutation.

We therefore aimed at assaying the potential self-sufficiency of this mutation to induce a more aggressive cancer phenotype. To this end, the functional impact of ectopically expressed G842C-mutated

PlxnB2 was compared to that of the WT receptor, in diverse tumor cells. The variant plexin proved capable to promote cell proliferation and tumor growth, as well as cancer invasiveness, consistent with the selective advantage of its acquisition and maintenance in CUPs. Although initially reported to bind *Sema4C* with high affinity, PlexinB2 is now considered a relatively promiscuous receptor of class-4 semaphorins, which probably explains its pivotal role in embryo development. In addition to semaphorins, PlexinB2 was reported to act as receptor for the unrelated molecule *angiogenin*, controlling cancer cell survival and growth, beyond angiogenesis (Yu *et al*, 2017). Moreover, it was recently reported that PlexinB2 acts as a cell surface mechanotransducer, adjusting cell adhesion and migration dynamics to matrix stiffness, and empowering glioblastoma cell invasiveness (Junqueira Alves *et al*, 2021). It is therefore possible to envisage a semaphorin-independent signaling competence of G842C-PlxnB2. Intriguingly, previous studies have reported that a complex conformational mechanism maintains the plexins in the inactive state, since IPT2-*sema* domain intermolecular interaction within dimeric plexin complexes results in receptor auto-inhibition, in the absence of the ligands (Marita *et al*, 2015; Kong *et al*, 2016). Unfortunately, the time course analyzable in our molecular modeling experiments on mutated receptors is limited to 500 ns of simulation, which is insufficient to reveal this kind of large conformational changes. Nevertheless, considering the conservation of the IPT3 domain in the plexin family (Junqueira Alves *et al*, 2019), it may be speculated that a mutation altering the stability of this highly structured domain may impact also on the state of the neighboring IPT2 and possibly destabilize the auto-inhibitory intermolecular complex. Notably, two additional CUP cases in our panel were found to carry *PLXNB2* mutations (summing up to 21% global incidence), though they did not concern Gly842 residue or the IPT3 domain. We analyzed the impact on receptor conformation and activity of these other genetic changes, as well as of additional mutations in the PlxnB2 IPT3 domain reported in non-CUP human tumors. However, none of the mutants was found to produce comparable effects as G842C, or in the presence of some mutations the activity seemed to be reduced, altogether supporting the signaling specificity of this novel variant found in CUPs. Future studies will tell if additional genetic variants of this receptor can promote oncogenic signaling in CUPs or other tumor types.

With the aim to elucidate the signaling cascade elicited downstream G842C-PlxnB2 mutant receptor, we posited the involvement of a tyrosine kinase receptor, since members of this superfamily have been previously associated with plexin signaling in cancer cells. Eventually, we identified EGFR as a previously unknown partner of PlxnB2 activity. In fact, PlxnB2 was found in complex with EGFR, and EGFR phosphorylation was enhanced in the presence of G842C-PlxnB2. Moreover, the greater invasiveness of CUP cells driven by the expression of the mutated plexin was abrogated by selective EGFR inhibitors applied in clinical practice, cetuximab and erlotinib.

In conclusion, we found that a novel activating mutation of the semaphorin receptor PlxnB2 sustains the proliferative autonomy of CUP stem cells and enhances their tumorigenic capacity and EGFR-dependent invasiveness. These data provide proof-of-principle of the functional involvement of an unexpected aberrant signaling pathway in CUP development and prompt for the characterization of additional axon guidance genes that we found to be mutated in human CUPs.



## Materials and Methods

### Patient recruitment, diagnosis, and tissue sample collection

Patients were enrolled at Candiolo Cancer Institute within Agnostos Profiling protocol (no. 010-IRCC-10IIS-15) approved by the Institute Ethical Committee. Informed consensus was obtained from all patients and the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report. The diagnostic workflow of CUPs followed ESMO guidelines (Fizazi *et al*, 2015). Fresh human specimens were collected and either stored in RNAlater (Life Technologies) or fixed in 4% buffered formaldehyde and embedded in paraffin. Sections were stained with hematoxylin and eosin to select tumor cell-rich areas before RNA extraction.

### RNA extraction, libraries preparation, and sequencing

Total RNA from FFPE from patient's tissues was purified using Maxwell<sup>®</sup> RSC RNA FFPE Kit (Promega). RNA was quantified using the Qubit 2.0 fluorimetric Assay (Thermo Fisher Scientific). Libraries were prepared from 100 ng of total RNA using the QuantSeq 3<sup>0</sup> mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen GmbH) and sequenced on a 15 NovaSeq 6000 sequencing system using an S1, 100 cycles flow cell (Illumina Inc.). Fastq files were generated using bcl2fastq (version v2.20.0.422, Illumina Inc.), and trimming was performed with bbdduk software (bbmap suite 20.37.31, Joint Genome Institute) and alignment on a human genome reference assembly (hg38) with STAR 2.6.0a (Dobin *et al*, 2013). Expression levels were determined with htseq-count 0.9.1 using cellRanger prebuild genes annotations (Single Cell Gene Expression, 10x Genomics; Ensembl Assembly 93). Data normalization was performed using edgeR (Anders *et al*, 2015).

### gDNA extraction, library preparation, and sequencing

Tumor gDNA was isolated using Relia PrepTM gDNA Tissue Miniprep System (Promega). Normal gDNA was derived from peripheral blood mononuclear cells (PBMCs) of the same patient using ReliaPrepTM Blood gDNA Miniprep System (Promega). DNA was quantified using Nanodrop ND1000 spectrophotometer (Thermo Fisher Scientific) and Qubit 4 Fluorometer (Thermo Fisher Scientific). Whole-exome sequencing with 150-bp paired reads was performed with a NextSeq 500 (Illumina) using 1  $\mu$ g genomic DNA and enrichment for whole exome performed according to SeqCap EZ MedExome (Roche). Adapters were clipped using Scythe (<https://github.com/vsbuffalo/scythe>) and trimmed using Sickel (<https://github.com/najoshi/sickle>). Alignment to the human genome (hg38) was done using Burrows–Wheeler Aligner (BWA) MEM (Li & Durbin, 2010). PCR duplicates were removed using rmdup of SAMtools (Li *et al*, 2009). Somatic SNVs and small insertion/deletions (InDels) were identified using Strelka2 (Kim *et al*, 2018). ANNOVAR (Wang *et al*, 2010) was used to annotate nonsilent (nonsynonymous, stopgain, stoploss, frameshift, nonframeshift, and splicing modifications) somatic mutations in each tumor.

### MEGA-V analysis

We used MEGA-V (Mutation Enrichment Gene set Analysis of Variants; Gambardella *et al*, 2017) tool to identify gene sets with a significantly higher number of variants in a CUP cohort compared with cohorts of patients affected by other cancer types. To this end, we first used TCGAAbiolinks (Colaprico *et al*, 2016) package in the R statistical environment to download from TCGA database the simple nucleotide variation datasets of somatic mutations in 33 distinct cancer types (*BRCA*, *AML*, *DLBC*, *CHOL*, *MESO*, *ACC*, *UCS*, *KICH*, *PCPG*, *ESCA*, *THYM*, *TGCT*, *UVM*, *CESC*, *BLCA*, *PAAD*, *LIHC*, *SKCM*, *UCEC*, *PRAD*, *THCA*, *OV*, *LGG*, *SARC*, *COAD*, *READ*, *KIRP*, *GBM*, *STAD*, *LUAD*, *KIRC*, *LUSC*, and *HNSC*), available at: <https://gdc.cancer.gov/>. We used MEGA-V tool to compare the frequency of mutations in 186 manually curated KEGG gene sets (<https://www.genome.jp/kegg/pathway.html>; Kanehisa & Goto, 2000) within samples of the CUP cohort, with respect to that observed in each of the 33 other cancer types found in TCGA dataset. Only nonsynonymous point mutations were considered for comparison (i.e., missense, nonsense, and nonstop mutations).

### Plexin-B2 systems preparation and molecular dynamics simulations

The human Plexin-B2 structure was retrieved from AlphaFold Database (Jumper *et al*, 2021; Varadi *et al*, 2022), available at <https://alphafold.ebi.ac.uk/>. In the present paper, only the Plexin-B2 extracellular domain was used; therefore, the whole structure was trimmed and the residues with higher confident score (local Distance Difference Test (pLDDT) > 50) have been kept; thus, the system comprehends the amino acidic sequence 21–1,190. In order to set up the p.G842C mutants, firstly Glycine 842 was substituted with a cysteine residue, using Maestro tools. This step was sufficient to generate the pseudo-wild-type mutant where the single point mutation did not perturb the sulfur bridge network, leaving the native bond between aa 845 and 860. Then, the disulfide bond between C845 and C860 was broken, and two different conformations were considered, respectively, between C842 and C845 (for the system p.G842C 842–845) and between C842 and C860 (for the system p.G842C 842–860). Finally, all four systems, the native protein, the pseudo-wild-type one and the two mutants with alternative disulfide bonds, have been minimized with Prime (Jacobson *et al*, 2002, 2004). Classical molecular dynamics simulations were performed with GROMACS 2020.4 package (Abraham *et al*, 2015), using the CHARMM36m force field (Huang *et al*, 2017) at full atomistic level using a TIP3 water solvent. The systems were solvated in a water box of dimension 13.5 × 12.9 × 15.12 Å under periodic boundary conditions. The total charge of the system was neutralized by randomly substituting water molecules with Na<sup>+</sup> ions and Cl<sup>-</sup> ions to obtain neutrality with 0.15 M salt concentration. Following a steepest descent minimization algorithm, the system was equilibrated in canonical ensemble (NVT) conditions for 125 ps, using Nose–Hoover thermostat with position restraints for the protein–peptide complexes. Immediately after this minimization procedure, all restraints were removed, and molecular dynamics runs were performed under NPT conditions at 303.15 K, using Nose–Hoover thermostat, with a T-coupling constant of 1 ps, and a Parrinello–Raman barostat at 1 atm. Van der Waals interactions



were modeled using 6–12 Lennard–Jones potential with a 1.2 nm cutoff. Long-range electrostatic interactions were calculated, with a cutoff for the real space term of 1.2 nm. All covalent bonds were constrained using the LINCS algorithm. The time step employed was 2 fs, and the coordinates were saved every 5 ps for analysis, which was performed using the MDAnalysis (Michaud-Agrawal *et al*, 2011) python library and PLUMED (Bonomi *et al*, 2009; Tribello *et al*, 2014; PLUMED consortium, 2019) tools.

#### BETARMSD analysis

PLUMED BETARMSD Collective Variable (CV) probes the antiparallel beta-sheet content of protein structures. Two protein segments containing three contiguous residues can form an antiparallel beta sheet. Although, if the two segments are part of the same protein chain, they must be separated by a minimum of two residues allowing enough space for the turn. This CV, thus, generates the set of all possible six residue sections that could form an antiparallel beta sheet and calculates the RMSD distance between the configuration in which the residues find themselves and an idealized antiparallel beta-sheet structure (Pietrucci & Laio, 2009). This is done by calculating the following sum of functions of the RMSD distances:

$$s \frac{1}{4} \sum \frac{1 - \frac{RMSD^n}{r_0^n}}{1 - \frac{RMSD^m}{r_0^m}}$$

where the sum runs over all possible segments of antiparallel beta sheet. For  $r_0$ , the default value of 0.8 was used.

#### Production of cDNA constructs expressing wild-type and mutated Plexin-B2

Expression constructs for wild-type human Plexin-B2 were reported previously (Gurrapu *et al*, 2018). In a plasmid encoding the full-length receptor, we replaced short cDNA restriction fragments with inserts containing *PLXNB2* mutations, produced through gene synthesis approach (supported by Biocat GmbH, Heidelberg, Germany). In addition to mutations identified in CUP samples in this study, genomic datasets of human non-CUP tumor samples found to carry other *PlxnB2* mutations in the IPT3 domain are available at the following links:

*R820H*: [www.cbioportal.org/patient=TCGA-EW-A1IZ](http://www.cbioportal.org/patient=TCGA-EW-A1IZ); [cancer.sanger.ac.uk/cosmic/sample/overview?id=1766751](http://cancer.sanger.ac.uk/cosmic/sample/overview?id=1766751);

*L828F*: [www.cbioportal.org/patient=TCGA-H4-A2HQ](http://www.cbioportal.org/patient=TCGA-H4-A2HQ);

*R843Q*: [www.cbioportal.org/patient=TCGA-B5-A3FC](http://www.cbioportal.org/patient=TCGA-B5-A3FC);

*Y852C*: [www.cbioportal.org/patient=TCGA-A5-A1OF](http://www.cbioportal.org/patient=TCGA-A5-A1OF); [cancer.sanger.ac.uk/cosmic/sample/overview?id=1759420](http://cancer.sanger.ac.uk/cosmic/sample/overview?id=1759420).

#### Cell culture

Agnospheres were isolated and cultured as described previously (Verginelli *et al*, 2021). Briefly, tumor samples were digested with collagenase I (Gibco), and after filtration, single-cell suspensions were resuspended in culture medium [Dulbecco's Modified Eagle's Medium: F12 (DMEM; Sigma) supplemented with N2 supplement (Life Technologies-GIBCO), BSA 0.5% (Sigma), Heparin 4  $\mu$ g/ml

(Sigma), 2 mM Glutamine (Sigma), and Penicillin–Streptomycin (EuroClone)]. Ultralow-attachment flasks (Corning, cat. CLS3814) were used in case of AS901. The same medium composition was used for further propagation of the agnospheres; when subculturing, agnospheres dissociation was mechanically achieved by pipetting and by trypsin treatment. MCF7 and HEK-293T cells were provided by American Type Culture Collection (ATCC) and cultured in DMEM or Iscove medium (Sigma-Aldrich), respectively, supplemented with 10% fetal bovine serum (FBS; Euroclone), 2 mM Glutamine (Sigma-Aldrich), and Penicillin–Streptomycin (EuroClone). The cells were periodically checked for mycoplasma contamination.

#### Antibodies and chemicals

Anti-PlexinB2 antibody used for Western blotting analysis was purchased from Abcam (ab193355; dilution 1:500). EGFR antibody was provided by Enzo Life Sciences (ALX-804-064-C100; dil. 1:500), while anti-phospho-EGFR-specific antibodies (directed to p-Tyr1068; ab5644; dil. 1:500) were from Abcam. Anti-vinculin (V4505; dil. 1:1,000) and anti-VSV-G (clone P5D4; dil. 1:1,000) were provided by Sigma-Aldrich. Anti-phospho-p44/42-MAPK (Erk1/2; Thr202/Tyr204; #4370; dil. 1:1,000), anti-p44/42-MAPK (Erk1/2; 137F5; #4695; dil. 1:1,000), and anti-phospho-S6 ribosomal protein (Ser240/244; #2215; dil. 1:1,000) antibodies were from Cell Signaling.

#### Ligand binding and “collapsed” phenotype analysis by cellular immunostaining

Semaphorin-binding assays were performed as described previously (Tamagnone *et al*, 1999). Briefly, expression constructs encoding WT or mutated *PlxnB2* (containing a VSV tag), or a control vector, were transiently transfected in COS-7 cells, by DEAE-Dextran method. Two days after transfection with the receptors, the cells were incubated for 1 h with alkaline phosphatase-conjugated Sema4C; after rinsing, the cells were fixed and the attached ligand was revealed by incubation with NBT/BCIP substrate mix (Promega, cat. S3771).

For immunofluorescence analysis of cell phenotype, transfected COS-7 cells were seeded on glass coverslips; 2 days after transfection, they were fixed in 4% paraformaldehyde for 15 min, permeabilized with 0.1% Triton/phosphate-buffered saline (PBS) for 3 min at room temperature, and then blocked by 5% donkey serum for 30 min. The fixed cells were then incubated with primary antibodies for 1 h at room temperature, followed by incubation with the fluorochrome-conjugated secondary antibodies for 30 min at room temperature. F-actin was stained by using fluorescent-Phalloidin conjugates. Nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI). The coverslips were then washed and mounted on slides. The images were acquired with a confocal laser-scanning microscope (SP5-Leica-CLSM) and analyzed using LAS AF LITE 2.6.0 software (Leica Application Suite); cells with one diameter lower than 40  $\mu$ m were scored as collapsed.

#### Lentiviral-mediated shRNA or gene transfer

*PLXNB2* knockdown was commonly achieved by lentiviral-mediated transfer of a validated puromycin-selectable construct expressing a targeted shRNA and GFP marker (Origene, cat.

TL317033B; targeting seq. 5'-CCACTGGCTGTGGAGCCGAAGCAAGTCCT-3'). For validation experiments, we transferred an independent shRNA sequence carried by TRCNo000048188 clone (targeting seq. 5'-GCTCTACCAATACACGCAGAA-3'), provided by Sigma-Aldrich. For overexpression experiments, a cDNA construct encoding human PlxnB2 (VSV-tagged, provided by Jun Takagi, Osaka, Japan) was subcloned into the lentiviral expression construct pLVX. Moreover, a cDNA fragment containing the sequence encoding PLXNB2-G842C mutation was produced by gene synthesis (BioCat GmbH, Heidelberg, Germany) and replaced to the wild-type sequence, by restriction site-mediated recombination, in the expression construct.

Lentiviral-mediated gene transfer was performed as described previously (Follenzi & Naldini, 2002; Brown et al, 2020). Briefly, nonreplicating viral particles containing constructs expressing cDNA or shRNAs (or pGFP-C-shRNA Vector [Origene], as control) were produced in HEK-293 T packaging cells by the calcium phosphate precipitation method. The harvesting of viral particles was carried out 48 h after transfection: the conditioned medium was filtered and centrifuged at 19,500 rpm for 2 h to obtain concentrated viral suspensions. Host cells were then incubated with viral particle-containing media in the presence of 8 µg/ml polybrene at 37° (multiplicity of infection [moi] = 5); CUP cells were dissociated from agnospheres and incubated with viral particles in suspension. Gene-transduced cells were then selected by 0.5 µg/ml puromycin treatment.

#### Cell viability assay

Agnospheres (400 cells/well) were seeded into 96-well cluster plates (in case of AS901, ultralow-attachment type; Corning, cat. CLS3474). Cell viability was evaluated at 24, 48, and 72 h by CellTiter-Glo® Luminescent Cell Viability Assay (Promega), according to the manufacturer's recommendations by employing the VICTOR X Multilabel Plate Readers (Perkin Elmer).

#### Limiting dilution assay

CUP cell clonogenic capacity was determined by limiting dilution assays (Hu & Smyth, 2009). Briefly, dissociated cells from different agnospheres were plated into 96-well plates at three different densities, with the following scheme: 30 wells with 1 cell/well, 18 wells with 5 cells/well, and 12 wells with 10 cells/well. After 4 weeks in culture, the number of agnosphere-containing wells was determined by checking under microscope. Positive test was considered wells with spheres larger than 100 µm. Extreme limiting dilution analysis (ELDA) online software was used to calculate the cancer cell stem frequency and statistical significance, by chi-squared test (Hu & Smyth, 2009).

#### Cell migration and invasion assays

For wound healing assays, MCF7 cells were cultured until 80–90% confluent into 6-well plate with complete medium. Cell monolayers were scratched manually with a micropipette tip; any detached cells were removed by washing with PBS, and the wells were replenished with fresh medium supplemented with 1% FBS. Low magnification images were taken by Leica DM1400B microscope (Leica Microsystems) and then analyzed by ImageJ (NIH, Bethesda, MD, USA)

measuring the scratched cell-free area at the time of the scratch ( $T_0$ ) and after 72 h ( $T_{72}$ ); wound closure % was computed as  $1 - (\text{cell-free area at } T_{72} / T_0) * 100$ .

Individual MCF-7 cell migration assays were performed using Transwell chamber inserts with a porous polycarbonate membrane (8 µm pore size) (Corning Costar Incorporated, Corning, NY, USA). Briefly, the lower side of the filter was coated with 10 µg/ml fibronectin.  $10 \times 10^4$  cells were added in the upper chamber in serum-free medium and then allowed to migrate for 24 h through the filter toward the lower chamber with 10% FBS-containing medium. At the end of the experiment, the nonmigrated cells on the upper side of the filter were removed by a cotton swab, followed by fixing with 4% paraformaldehyde, and cell staining with crystal violet. The microscopic images were then quantified either by cell counting or by quantifying the integrated pixel values using ImageJ.

For invasion assays,  $1 \times 10^5$  AS906 cells were added to the upper chamber of a Transwell insert coated with Matrigel (40 µg; Cultrex Reduced Growth Factor Basement Membrane Extract, Type 2, Path-clear, R&D) and allowed to migrate in the lower chamber. After 48 h, the cells adherent to the lower side of the porous membrane were fixed and analyzed as above.

#### ADX preclinical model

Agnospheres-derived xenograft preclinical mouse model (ADX) has been described previously (Verginelli et al, 2021). Approx. eight-week-old female NOD-SCID mice were purchased from Charles River (Lecco, Italy), and 50,000 cells were injected subcutaneously in each mouse. Once tumor mass became palpable, the tumor burden was measured twice weekly, and the volume was estimated by the formula  $(a^2 \times b) \times 0.52$ , where a is the minor and b is the major tumor diameter. Mice were sacrificed 4 or 8 weeks after injection, for ADX901 and ADX43, respectively.

All animal procedures were approved by the Ethical Committee of the University of Turin (Candiolo, Turin, Italy) and by the Italian Ministry of Health (auth. n. 741/2020-PR) and were conducted in compliance with European laws and policies. Throughout the study, the mice were kept at a temperature of  $22 \pm 1^\circ$  C and a relative humidity of  $60 \pm 5\%$ , with a 12 h light/dark cycle and 12 air changes/h.

#### Western blotting analysis and protein immunoprecipitation

The analysis of phosphorylated proteins in CUP cells was done 3 h after cell resuspension at a density of 125,000 cells/ml, in fresh serum-free medium. Cells were lysed in Cell Lysis Buffer (Cell Signaling cat#9803) containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM Na<sub>2</sub>EDTA, 1 mM EGTA, 1% Triton, 2.5 mM sodium pyrophosphate, 1 mM beta-glycerophosphate, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 1 µg/ml leupeptin, protease inhibitor cocktail (Sigma-Aldrich). Cellular lysates were incubated for 30 min on ice and then centrifuged at 15,000 g, 15 min, at 4° C. Protein concentration of cell extracts was determined by using Pierce bicinchoninic acid (BCA) reagent (Thermo Fisher Scientific, cat. 23227) according to the manufacturer's instructions. Protein samples were denatured by adding a 4× loading buffer (β-mercaptoethanol 0.6 mol/l; SDS 8%; Tris-HCl 0.25 mol/l pH 6.8; glycerol 40%; bromophenol blue 0.2%), incubated at 95° C for 5 min. Samples containing equivalent amounts of

protein were subjected to 7.5% SDS–PAGE. Proteins were transferred onto a nitrocellulose membrane using the Trans-Blot Turbo Transfer System (Bio-Rad) according to the manufacturer's instructions, probed with Abs of interest, and revealed by enhanced chemiluminescence technique, using Chemidoc Image Lab analyzer and software (Bio-Rad), according to the manufacturer protocols.

For protein immunoprecipitation experiments, HEK-293T cells were transiently transfected to express either PlexinB2-WT or PlexinB2-G842C, in association with EGFR. 48 h after the transient transfection, the cells were transferred on ice, washed three times in cold PBS 1×, and lysed with a buffer containing 20 mM Tris–HCl (pH 7.5), 150 mM NaCl, 1 mM Na<sub>2</sub>EDTA, 1 mM EGTA, 1% Triton, 2.5 mM sodium pyrophosphate, 1 mM beta-glycerophosphate, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 1 μg/ml leupeptin, protease inhibitor cocktail (Sigma-Aldrich). Cellular lysates were incubated for 30 min on ice and then centrifuged at 15,000 g, 15 min, at 4° C. The total protein amount was determined using the bicinchoninic acid protein assay reagent (Thermo Fisher Scientific). Equivalent amounts (1 mg) of total proteins were incubated with anti-PlexinB2 antibodies and Dynabeads Protein G (Invitrogen, 10004D), for 4 h at 4° C. Immuno-complexes were washed four times with lysis buffer and then separated by SDS–PAGE. Proteins were finally transferred to a nitrocellulose membrane (Bio-Rad) and analyzed as described previously.

#### Data analysis and statistics

In general, blinding was not applied, but data analysis was validated by at least two scientists.

Results of experiments in which appropriate negative/positive controls were not validated were consequently excluded from the analysis. The statistical significance of quantitative data was analyzed by GraphPad Prism 8.0.0 software, applying the most appropriate methods and correction tests, specified in individual figure legends. Exact calculated *P*-values are also indicated in the respective figure legends, apart in case of low-end values: \*\*\*\**P* < 0.0001.

#### Data availability

Datasets containing WES raw data of CUP samples described in this study are available at the European Genome–Phenome Archive (EGA; <https://ega-archive.org/studies/>), under the accession codes EGAS00001006621, EGAS00001004868, and EGAS00001004059. Raw data relative to RNA-Seq are available from the Gene Expression Omnibus (GEO) repository, under the accession code GSE167473.

Expanded View for this article is available [online](#).

#### Acknowledgments

The authors are grateful to Prof. Raffaele Calogero (University of Torino) for advice in gene expression analysis, to Prof. Mattia Lauriola (University of Bologna) for advice in EGFR signaling analysis, and to Silvia Benvenuti, Melissa Milan, and Giorgia Innamorati for help and advice in certain experiments. This work was supported by AIRC—Italian Association for Cancer Research (Molecular Clinical Oncology 5 × 1000 Grant #21052, to P.M.C., C.B., E.G., and L.T.;

#### The paper explained

##### Problem

Cancer of unknown primary (CUP) is a pathological entity represented by metastatic tumors that are first time diagnosed in the absence of clinically detectable primary lesions. The genetic makeup of these tumors, and the mechanisms underlying their accelerated metastatic dissemination, are still elusive. Hence, currently applied therapies are neither specific nor usually effective. Therefore, a better definition of CUP genetic signature, and the identification of new suitable molecular therapeutic targets, may lead to the design of innovative effective treatments, increasing patient survival.

##### Results

The genomic analysis of confirmed CUP patients revealed a higher frequency of mutations in axon guidance genes, compared with other tumor types. In this study, we particularly focused on a novel activating mutation of PlexinB2 (G842C-PlxnB2), capable of inducing proliferation and invasiveness of CUP stem cells. Notably, the depletion of G842-PlxnB2 resulted in the impairment of cancer cell self-renewal and tumorigenic growth in mice. On the contrary, the overexpression of mutated PlexinB2 enhanced growth and tumorigenic capacity in mice. Remarkably, we found that the sustained proliferative autonomy and invasive properties induced by G842C-PlxnB2 in CUP cells are dependent on the oncogenic tyrosine kinase epidermal growth factor receptor (EGFR) and were abrogated by the treatment with selective EGFR inhibitors validated for clinical use.

##### Impact

The finding of mutations in axon guidance genes, PLXNB2 in particular, may provide novel genetic biomarkers guiding CUP disease management. Moreover, the identification of EGFR signaling deregulation, as functional consequence of the aberrant activation of axon guidance pathways, may prompt the development of novel therapeutic approaches for CUP patients carrying the implicated mutations.

moreover, V.N. was supported by an AIRC Fellowship for Italy), by the Italian Ministry for Research (PRIN-2017TATYMP grant, to L.T.), and by the Italian Ministry of Health (Ricerca Corrente grants 2022 to E.G. [RRC-2022-23680791]; and to L.T. [FPG-IRCCS]). Università Cattolica del Sacro Cuore also contributed to the funding of this research project and to its publication. This work was furthermore supported by the CINECA supercomputing centers through the grant IsC97 (n.HP10CKTQ6H).

#### Author contributions

Serena Brundu: Validation; investigation; visualization; writing—review and editing. Virginia Napolitano: Validation; investigation; visualization; writing—review and editing. Giulia Franzolin: Validation; investigation; visualization. Ettore Lo Cascio: Investigation; visualization. Roberta Mastrantonio: Validation; investigation; visualization. Gabriele Sardo: Validation; investigation; visualization. Eliano Cascardi: Investigation; methodology. Federica Verginelli: Resources; validation; investigation; methodology. Sergio Sarnataro: Software; validation; investigation; methodology. Gennaro Gambardella: Data curation; software; supervision; validation. Alberto Pisacane: Data curation; supervision; methodology. Alessandro Arcovito: Conceptualization; data curation; software; supervision; validation; writing—review and editing. Carla Boccaccio: Conceptualization; resources; data curation; writing—review and editing. Paolo M Comoglio: Conceptualization; supervision; funding acquisition; project administration. Enrico Girardo: Conceptualization; supervision; funding acquisition; writing—original draft; project administration; writing—review and editing.

Luca Tamagnone: Conceptualization; supervision; funding acquisition; visualization; writing—original draft; project administration; writing—review and editing.

## Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

## References

- Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2: 19–25
- Anders S, Pyl PT, Huber W (2015) HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166–169
- Artigiani S, Barberis D, Fazzari P, Longati P, Angelini P, van de Loo J-W, Comoglio PM, Tamagnone L (2003) Functional regulation of semaphorin receptors by proprotein convertases. *J Biol Chem* 278: 10094–10101
- Benvenuti S, Milan M, Geuna E, Pisacane A, Senetta R, Gambardella G, Stella GM, Montemurro F, Sapino A, Boccaccio C *et al* (2020) Cancer of unknown primary (CUP): genetic evidence for a novel nosological entity? A case report. *EMBO Mol Med* 12: e11756
- Biankin AV, Waddell N, Kassahn KS, Gingras M-C, Muthuswamy LB, Johns AL, Miller DK, Wilson PJ, Patch A-M, Wu J *et al* (2012) Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* 491: 399–405
- Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, Donadio D, Marinelli F, Pietrucci F, Broglia RA *et al* (2009) PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun* 180: 1961–1972
- Brown LY, Dong W, Kantor B (2020) An improved protocol for the production of lentiviral vectors. *STAR Protoc* 1: 100152
- Cagnoni G, Tamagnone L (2014) Semaphorin receptors meet receptor tyrosine kinases on the way of tumor progression. *Oncogene* 33: 4795–4802
- Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, Davidson S, Edwards M, Elvin JA, Hodel KP *et al* (2017) Comprehensive analysis of hypermutation in human cancer. *Cell* 171: 1042–1056.e10
- Chedotal A, Kerjan G, Moreau-Fauvarque C (2005) The brain within the tumor: new roles for axon guidance molecules in cancers. *Cell Death Differ* 12: 1044–1056
- Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I *et al* (2016) TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 44: e71
- Deng S, Hirschberg A, Worzfeld T, Penachioni JY, Korostylev A, Swiercz JM, Vodrazka P, Mauti O, Stoeckli ET, Tamagnone L *et al* (2007) Plexin-B2, but not plexin-B1, critically modulates neuronal migration and patterning of the developing nervous system *in vivo*. *J Neurosci* 27: 6333–6347
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21
- Fard D, Tamagnone L (2021) Semaphorins in health and disease. *Cytokine Growth Factor Rev* 57: 55–63
- Fizazi K, Greco FA, Pavlidis N, Daugaard G, Oien K, Pentheroudakis G, ESMO Guidelines Committee (2015) Cancers of unknown primary site: ESMO clinical practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 26: v133–v138
- Follenzi A, Naldini L (2002) HIV-based vectors. Preparation and use. *Methods Mol Med* 69: 259–274
- Gambardella G, Cereda M, Benedetti L, Ciccarelli FD (2017) MEGA-V: detection of variant gene sets in patient cohorts. *Bioinformatics* 33: 1248–1249
- Gurrapu S, Pupo E, Franzolin G, Lanzetti L, Tamagnone L (2018) Sema4C/PlexinB2 signaling controls breast cancer cell growth, hormonal dependence and tumorigenic potential. *Cell Death Differ* 25: 1259–1275
- Hinck L (2004) The versatile roles of «axon guidance» cues in tissue morphogenesis. *Dev Cell* 7: 783–793
- Hu Y, Smyth GK (2009) ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J Immunol Methods* 347: 70–78
- Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 14: 71–73
- Huang Y, Tejero R, Lee VK, Brusco C, Hannah T, Bertucci TB, Junqueira Alves C, Katsyiv I, Kluge M, Foty R *et al* (2021) Plexin-B2 facilitates glioblastoma infiltration by modulating cell biomechanics. *Commun Biol* 4: 145
- Jacobson MP, Friesner RA, Xiang Z, Honig B (2002) On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol* 320: 597–608
- Jacobson MP, Pincus DL, Rapp CS, Day T, Honig B, Shaw DE, Friesner RA (2004) A hierarchical approach to all-atom protein loop prediction. *Proteins* 55: 351–367
- Jiang C, Javed A, Kaiser L, Nava MM, Xu R, Brandt DT, Zhao D, Mayer B, Fernandez-Baldovinos J, Zhou L *et al* (2021) Mechanochemical control of epidermal stem cell divisions by B-plexins. *Nat Commun* 12: 1308
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A *et al* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583–589
- Junqueira Alves C, Dariolli R, Haydak J, Kang S, Hannah T, Wiener RJ, DeFronzo S, Tejero R, Gusella GL, Ramakrishnan A *et al* (2021) Plexin-B2 orchestrates collective stem cell dynamics via actomyosin contractility, cytoskeletal tension and adhesion. *Nat Commun* 12: 6019
- Junqueira Alves C, Yotoko K, Zou H, Friedel RH (2019) Origin and evolution of plexins, semaphorins, and met receptor tyrosine kinases. *Sci Rep* 9: 1970
- Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27–30
- Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, Chen X, Kim Y, Beyter D, Krusche P *et al* (2018) Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* 15: 591–594
- Kong Y, Janssen BJC, Malinauskas T, Vangoor VR, Coles CH, Kaufmann R, Ni T, Gilbert RJC, Padilla-Parra S, Pasterkamp RJ *et al* (2016) Structural basis for plexin activation and regulation. *Neuron* 91: 548–560
- Le AP, Huang Y, Pingle SC, Kesari S, Wang H, Yong RL, Zou H, Friedel RH (2015) Plexin-B2 promotes invasive growth of malignant glioma. *Oncotarget* 6: 7293–7304
- Li H, Durbin R (2010) Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics (Oxford, England)* 26: 589–595
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079
- Li Y, Ohira M, Zhou Y, Xiong T, Luo W, Yang C, Li X, Gao Z, Zhou R, Nakamura Y *et al* (2017) Genomic analysis-integrated whole-exome sequencing of neuroblastomas identifies genetic mutations in axon guidance pathway. *Oncotarget* 8: 56684–56697



- Löffler H, Pfarr N, Kriegsmann M, Endris V, Hielscher T, Lohneis P, Folprecht G, Stenzinger A, Dietel M, Weichert W *et al* (2016) Molecular driver alterations and their clinical relevance in cancer of unknown primary site. *Oncotarget* 7: 44322–44329
- Ma Q, Zhang K, Yao H-P, Zhou Y-Q, Padhye S, Wang M-H (2010) Inhibition of MSP-RON signaling pathway in cancer cells by a novel soluble form of RON comprising the entire sema sequence. *Int J Oncol* 36: 1551–1561
- Marita M, Wang Y, Kaliszewski MJ, Skinner KC, Comar WD, Shi X, Dasari P, Zhang X, Smith AW (2015) Class A plexins are organized as preformed inactive dimers on the cell surface. *Biophys J* 109: 1937–1945
- Mehta V, Pang K-L, Rozbesky D, Nather K, Keen A, Lachowski D, Kong Y, Karia D, Ameismeier M, Huang J *et al* (2020) The guidance receptor plexin D1 is a mechanosensor in endothelial cells. *Nature* 578: 290–295
- Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* 32: 2319–2327
- Navis AC, van Lith SAM, van Duijnhoven SMJ, de Pooter M, Yetkin-Arik B, Wesseling P, Hendriks WJA, Venselaar H, Timmer M, van Cleef P *et al* (2015) Identification of a novel MET mutation in high-grade glioma resulting in an auto-active intracellular protein. *Acta Neuropathol* 130: 131–144
- Neufeld G, Mumblat Y, Smolkin T, Toledano S, Nir-Zvi I, Ziv K, Kessler O (2016) The semaphorins and their receptors as modulators of tumor progression. *Drug Resist Updates* 29: 1–12
- Olivier T, Fernandez E, Labidi-Galy I, Dietrich P-Y, Rodriguez-Bravo V, Baciarello G, Fizazi K, Patrikidou A (2021) Redefining cancer of unknown primary: is precision medicine really shifting the paradigm? *Cancer Treat Rev* 97: 102204
- Pavlidis N, Pentheroudakis G (2012) Cancer of unknown primary site. *Lancet* 379: 1428–1435
- Pietrucci F, Laio A (2009) A collective variable for the efficient exploration of protein Beta-sheet structures: application to SH3 and GB1. *J Chem Theory Comput* 5: 2197–2201
- Pisacane A, Cascardi E, Berrino E, Polidori A, Sarotto I, Casorzo L, Panero M, Boccaccio C, Verginelli F, Benvenuti S *et al* (2022) Real-world histopathological approach to malignancy of undefined primary origin (MUO) to diagnose cancers of unknown primary (CUPs). *Virchows Arch* 8: 1–13
- PLUMED Consortium (2019) Promoting transparency and reproducibility in enhanced molecular simulations. *Nat Methods* 16: 670–673
- Rassy E, Pavlidis N (2020) Progress in refining the clinical management of cancer of unknown primary in the molecular era. *Nat Rev Clin Oncol* 17: 541–554
- Ross JS, Sokol ES, Moch H, Mileskin L, Baciarello G, Losa F, Beringer A, Thomas M, Elvin JA, Ngo N *et al* (2021) Comprehensive genomic profiling of carcinoma of unknown primary origin: retrospective molecular classification considering the CUPISCO study design. *Oncologist* 26: e394–e402
- Ross JS, Wang K, Gay L, Otto GA, White E, Iwanik K, Palmer G, Yelensky R, Lipson DM, Chmielecki J *et al* (2015) Comprehensive genomic profiling of carcinoma of unknown primary site: new routes to targeted therapies. *JAMA Oncol* 1: 40–49
- Sigismund S, Avanzato D, Lanzetti L (2018) Emerging functions of the EGFR in cancer. *Mol Oncol* 12: 3–20
- Stella GM, Benvenuti S, Gramaglia D, Scarpa A, Tomezzoli A, Cassoni P, Senetta R, Venesio T, Pozzi E, Bardelli A *et al* (2011) MET mutations in cancers of unknown primary origin (CUPs). *Hum Mutat* 32: 44–50
- Subbiah IM, Tsimberidou A, Subbiah V, Janku F, Roy-Chowdhuri S, Hong DS (2017) Next generation sequencing of carcinoma of unknown primary reveals novel combinatorial strategies in a heterogeneous mutational landscape. *Oncotargets Ther* 4: 47–56
- Tamagnone L, Artigiani S, Chen H, He Z, Ming GI, Song H, Chedotal A, Winberg ML, Goodman CS, Poo M *et al* (1999) Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. *Cell* 99: 71–80
- Tan H, Bao J, Zhou X (2015) Genome-wide mutational spectra analysis reveals significant cancer-specific heterogeneity. *Sci Rep* 5: 12566
- Tothill RW, Li J, Mileskin L, Doig K, Siganakis T, Cowin P, Fellowes A, Semple T, Fox S, Byron K *et al* (2013) Massively-parallel sequencing assists the diagnosis and guided treatment of cancers of unknown primary. *J Pathol* 231: 413–423
- Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) PLUMED 2: new feathers for an old bird. *Comput Phys Commun* 185: 604–613
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A *et al* (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 50: D439–D444
- Varghese AM, Arora A, Capanu M, Camacho N, Won HH, Zehir A, Gao J, Chakravarty D, Schultz N, Klimstra DS *et al* (2017) Clinical and molecular characterization of patients with cancer of unknown primary in the modern era. *Ann Oncol* 28: 3015–3021
- Verginelli F, Pisacane A, Gambardella G, D'Ambrosio A, Candiello E, Ferrio M, Panero M, Casorzo L, Benvenuti S, Cascardi E *et al* (2021) Cancer of unknown primary stem-like cells model multi-organ metastasis and unveil liability to MEK inhibition. *Nat Commun* 12: 2498
- Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38: e164
- Yang H, Yuan L, Ibaragi S, Li S, Shapiro R, Vanli N, Goncalves KA, Yu W, Kishikawa H, Jiang Y *et al* (2022) Angiogenin and plexin-B2 axis promotes glioblastoma progression by enhancing invasion, vascular association, proliferation and survival. *Br J Cancer* 127: 422–435
- Yazdani U, Terman JR (2006) The semaphorins. *Genome Biol* 7: 211
- Yu W, Goncalves KA, Li S, Kishikawa H, Sun G, Yang H, Vanli N, Wu Y, Jiang Y, Hu MG *et al* (2017) Plexin-B2 mediates physiologic and pathologic functions of angiogenin. *Cell* 171: 849–864.e25
- Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, Srinivasan P, Gao J, Chakravarty D, Devlin SM *et al* (2017) Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* 23: 703–713



License: This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



## Expanded View Figures

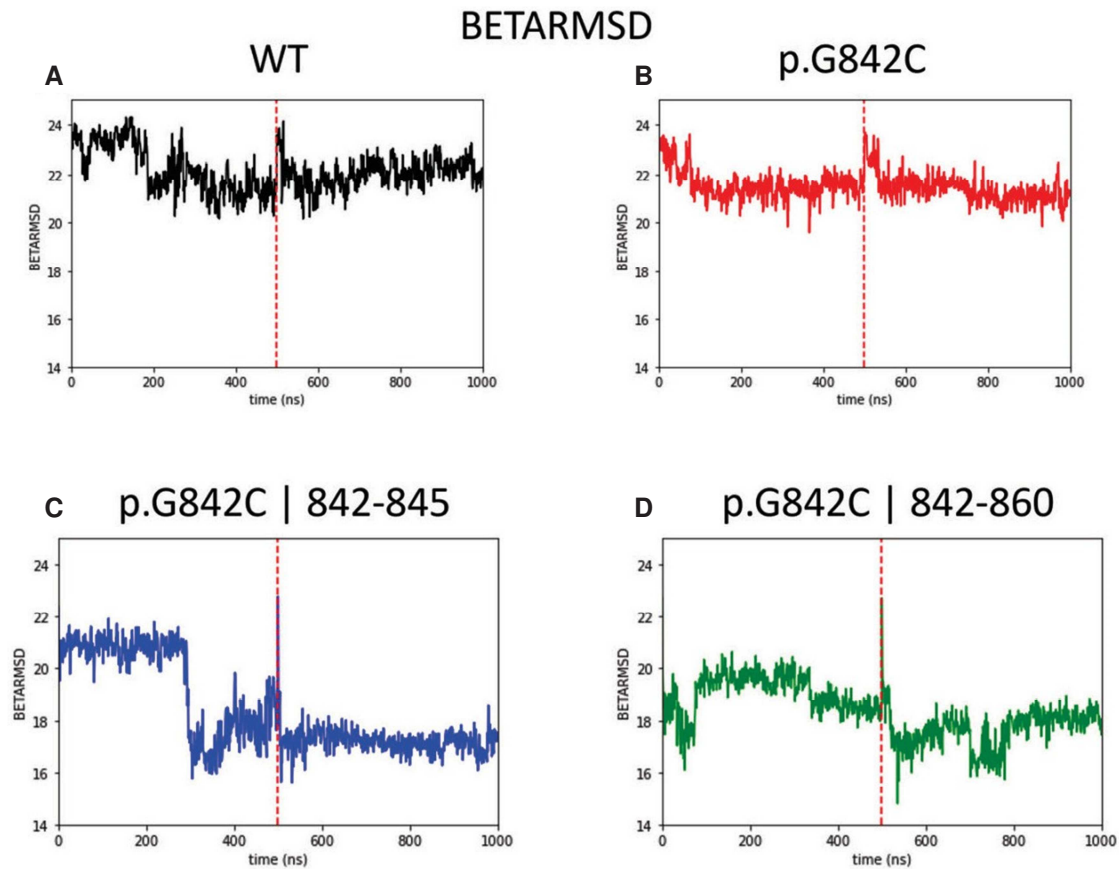


Figure EV1. BETARMSD plots of WT PlxnB2 and G842C-mutated isoforms.

A–D BETARMSD plots of the concatenated replicas of PlxnB2 WT, p.G842C, p.G842C 842–845, and p.G842C 842–860, respectively. The plots report the beta-sheet content evolution among the four studied systems, using the BETARMSD score described in the [Materials and Methods](#) section. In panel A, it is reported the time course of the selected parameter for the two concatenated MD simulations of the wild-type protein. It is evident that, starting from the initial state, a small rearrangement of the secondary structure is present even if the parameter remains quite stable along the simulation as identified by the average value obtained that is  $\langle \text{BETARMSD}_{\text{WT}} \rangle = 21.9 \pm 0.5$ . The behavior of the same parameter for the point mutation of G842 in a cysteine residue, still maintaining the native disulfide bond, is reported in panel B, and it shows a slight decrease of beta-sheet content ( $\langle \text{BETARMSD}_{\text{p.G842C}} \rangle = 20.4 \pm 0.4$ ), compared with WT due to the loss of a single beta strand ( $\beta\text{G}$ , as it is shown in Fig 3A in the main text). The same point mutation coupled with alternative disulfide bonds leads to a more relevant loss of beta-sheet content in both cases ( $\langle \text{BETARMSD}_{\text{p.G842C}\{842-845\}} \rangle = 17.6 \pm 1.19$ ,  $\langle \text{BETARMSD}_{\text{p.G842C}\{842-860\}} \rangle = 18.4 \pm 0.80$ ) due to the disruption of  $\beta\text{G}$ ,  $\beta\text{C}$ , and  $\beta\text{D}$  strands (see Fig 3D and E in the main text).

Figure EV2. RMSD plots of additional PlxnB2 ectodomain mutants found in human tumors.

A–D RMSD plots of two independent repetitions of 250 ns molecular dynamics of diverse PlxnB2 IPT3 domain mutants found in human tumors: (A) Mutant R820H; (B) mutant L828F; (C) mutant R843Q; (D) mutant Y852C.  
 E, F RMSD plots of two independent repetitions of 250 ns molecular dynamics of additional PlxnB2 mutants (outside IPT3 domain) found in CUP samples: (E) R531P and (F) P1058S.  
 G, H RMSD plots of two independent repetitions of 500 ns molecular dynamics of wild-type PlxnB2 and G842C mutant, extracted from Fig 3 and reported here for internal reference.

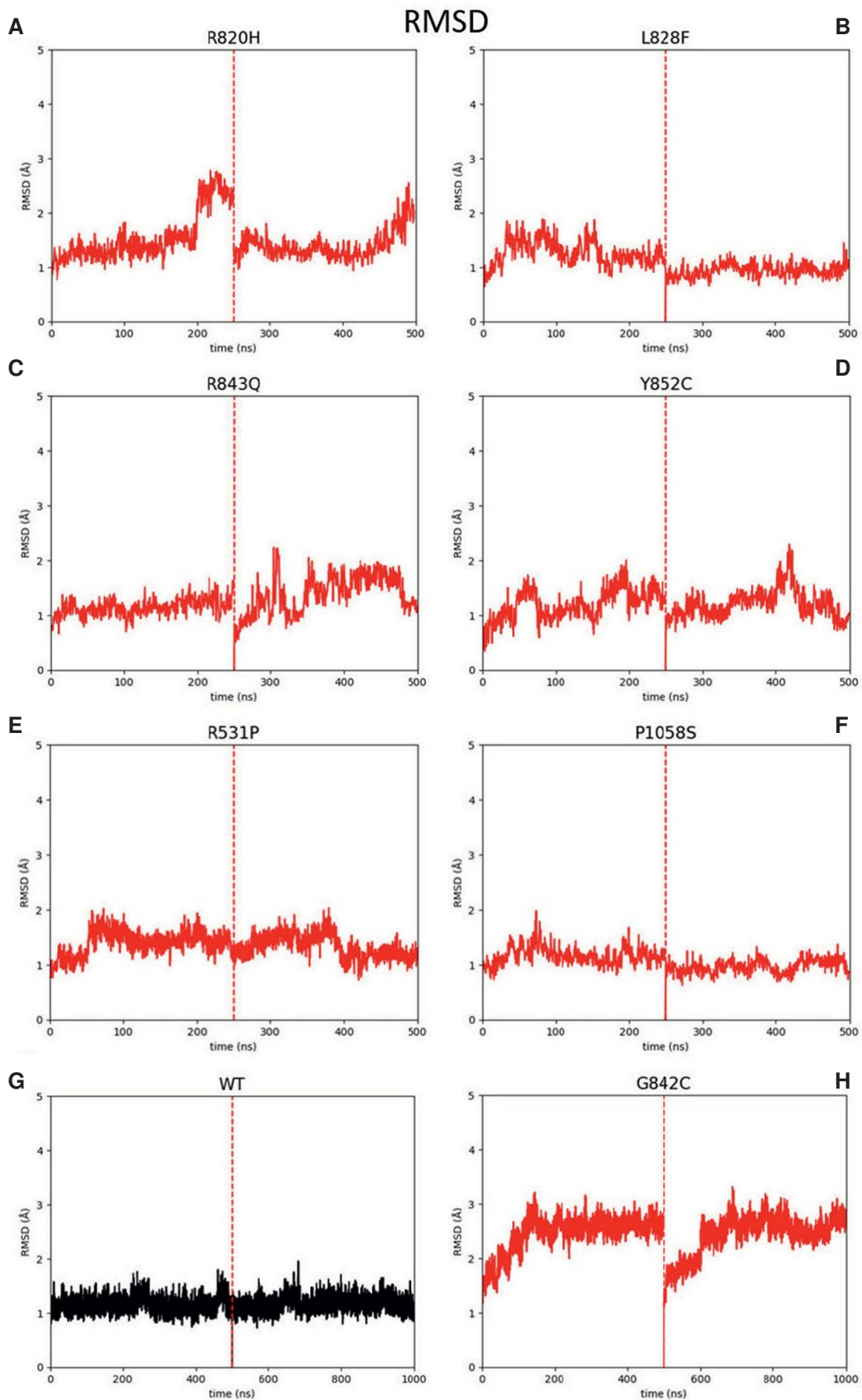


Figure EV2.

Figure EV3. Validation of CUP models with modified expression of PlxnB2 WT or G842C mutant.

- A Western blotting analysis of PlxnB2 expression in the indicated CUP-derived agnospheres, either subjected to gene knockdown by lentiviral-mediated transfer of targeted shRNAs (shPlxnB2) or transduced with a nontargeting control sequence (shScramble).
- B Western blotting analysis of PlxnB2 expression in CUP-derived AS67, engineered by lentiviral-mediated transfer of an empty vector control (mock), nontargeting scrambled shRNAs, PlxnB2-targeted shRNAs, a wild-type PlxnB2 expression construct, or a G842C-mutated PlxnB2 expression construct PlxnB2-G842C, respectively.
- C Western blotting analysis of PlxnB2 expression in AS901 and AS906 (as indicated), engineered to overexpress either wild-type PlxnB2 or PlxnB2-G842C, or transduced with an empty vector (mock).
- D Western blotting analysis of PlxnB2 expression in AS43 subjected to gene knockdown by lentiviral-mediated transfer of an alternative independent shRNA sequence (shPlxnB2 #2, see [Materials and Methods](#)), or transduced with a corresponding empty vector (shCtrl).
- E Time course analysis of cellular viability in shPlxnB2-transduced and control agnospheres described in the previous panel, over 7 days of growth in culture. Values are mean  $\pm$  SD of  $n = 3$  independent experiments (with quadruplicate technical replicates). The statistical significance was assessed by 2-way ANOVA multiple comparisons for each time point, with Bonferroni correction. shCtrl vs. shPlxnB2#2(1): at day 3  $**P = 0.0086$ , day 4  $**P = 0.0047$ ;  $****P < 0.0001$ . shCtrl vs. shPlxnB2#2(2):  $§P = 0.0161$ ;  $P < 0.0001$ .

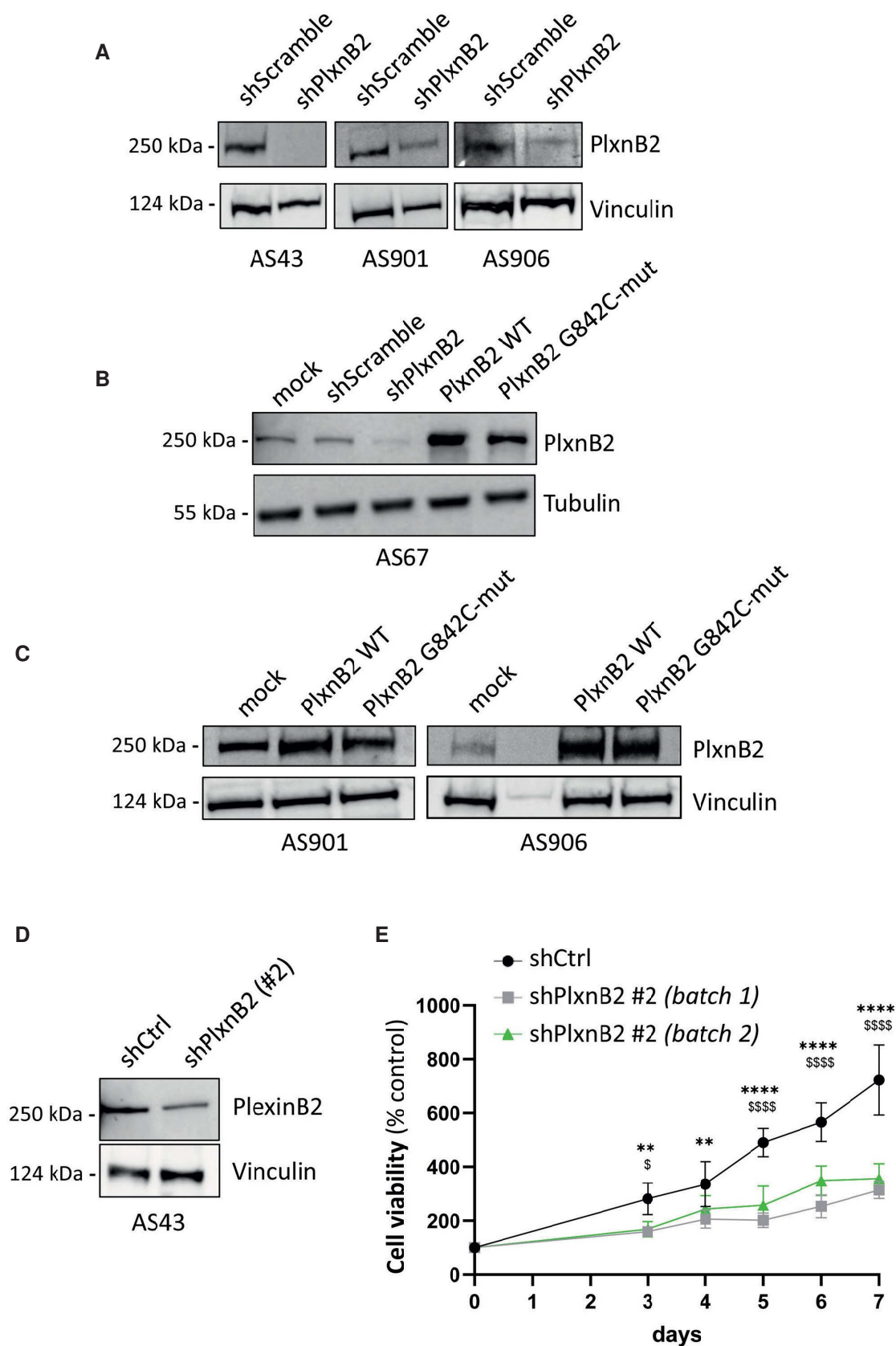


Figure EV3.

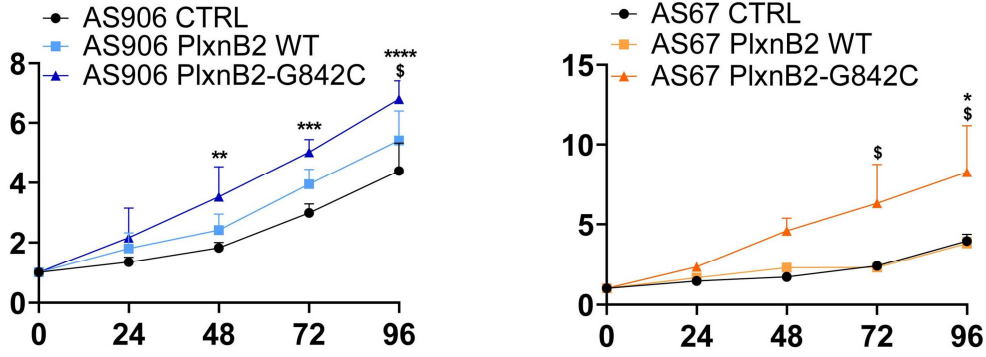


Figure EV4. Differential regulation of CUP cell growth by WT PlxnB2 or G842C mutant.

Time course cell viability analysis of AS906 (on the left) and AS67 (on the right) overexpressing either wild-type or PlxnB2-G842C, or mock controls transduced with an empty vector. Plotted values are the mean  $\pm$  SD of  $n = 3$  independent experiments (six technical replicates for each). The statistical significance was assessed by two-way ANOVA test. AS906 PlxnB2-G842C vs. controls (CTRL): at 48 h  $**P = 0.0042$ , 72 h  $***P = 0.0007$ , 96 h  $****P < 0.0001$ . AS906 PlxnB2-G842C vs. PlxnB2-WT: at 96 h  $^{\$}P = 0.0224$ . AS67 PlxnB2-G842C vs. controls (CTRL): at 96 h  $*P = 0.0337$ . AS67 PlxnB2-G842C vs. PlxnB2-WT: at 72 h  $^{\$}P = 0.0494$ , 96 h  $^{\$}P = 0.0274$ .



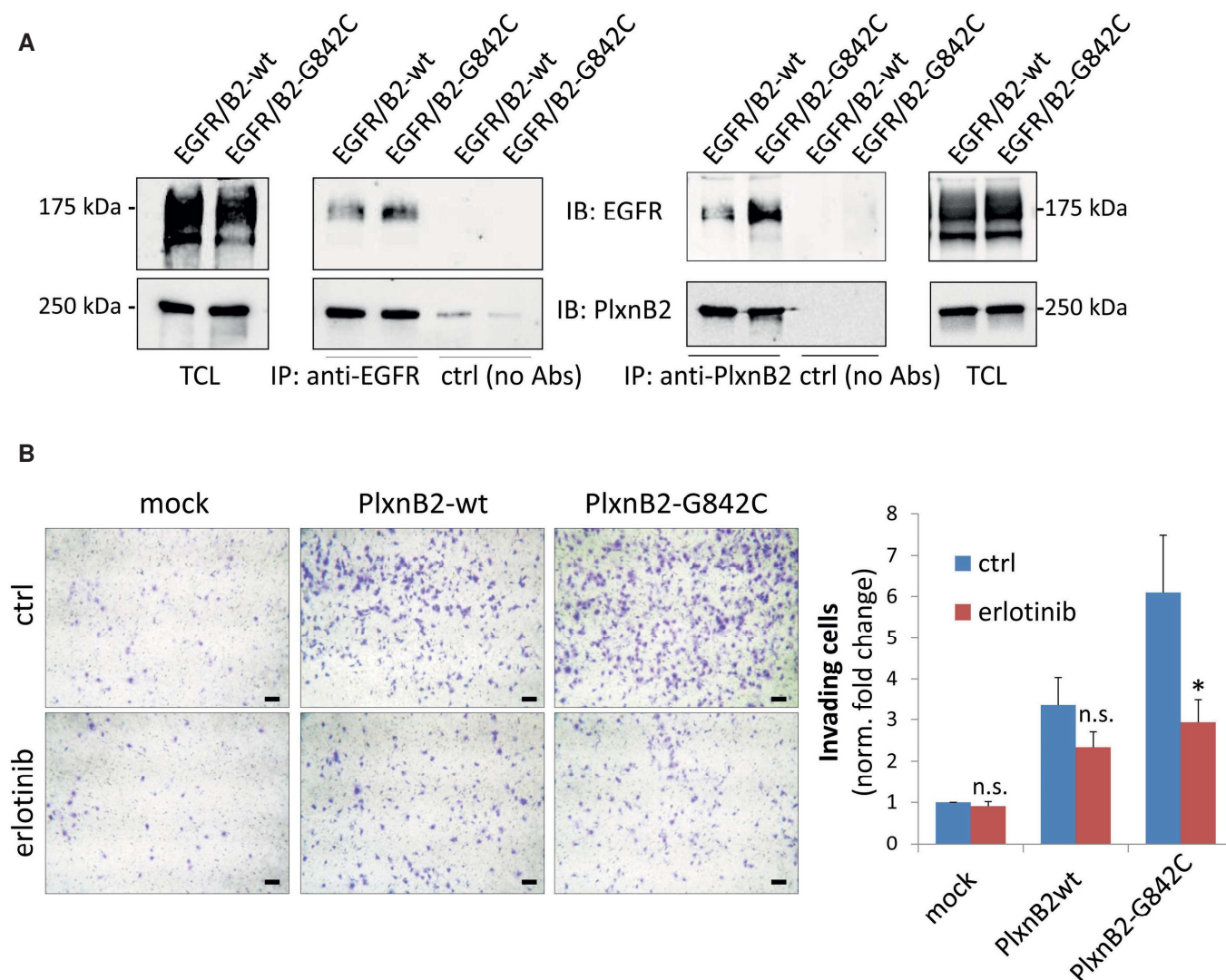


Figure EV5. PlxnB2-EGFR complex and EGFR-dependent regulation of CUP cell invasiveness.

A HEK293T cells were co-transfected with EGFR and either wild-type or PlxnB2-G842C. On the left, total cell lysates (TCL) were subjected (or not) to immunoprecipitation with anti-EGFR, and immunoblotted with the indicated antibodies. Data shown are representative of  $n = 3$  experiments. On the right, EGFR-PlxnB2 co-immunoprecipitation was reciprocally assessed in HEK293T cells co-transfected as above, by subjecting total cell lysates to immunoprecipitation with anti-PlxnB2 antibodies. Data shown are representative of  $n = 2$  experiments.

B Similar to the experiment shown in main Fig 9D, the invasiveness of AS906 cells either mock, or overexpressing WT or PlxnB2-G842C, was assessed in matrigel-coated Transwell inserts, in the presence or absence of the EGFR inhibitor erlotinib  $1 \mu\text{M}$ . Invading cells were stained with crystal violet, photographed (see representative low magnification images on the left; scale bar:  $100 \mu\text{m}$ ), and quantified by ImageJ. Plotted values are the mean  $\pm$  SD of  $n = 3$  independent experiments. The statistical significance across replicates was verified by unpaired  $t$ -test multiple comparisons between erlotinib-treated and untreated conditions, per each group:  $*P = 0.021$ .

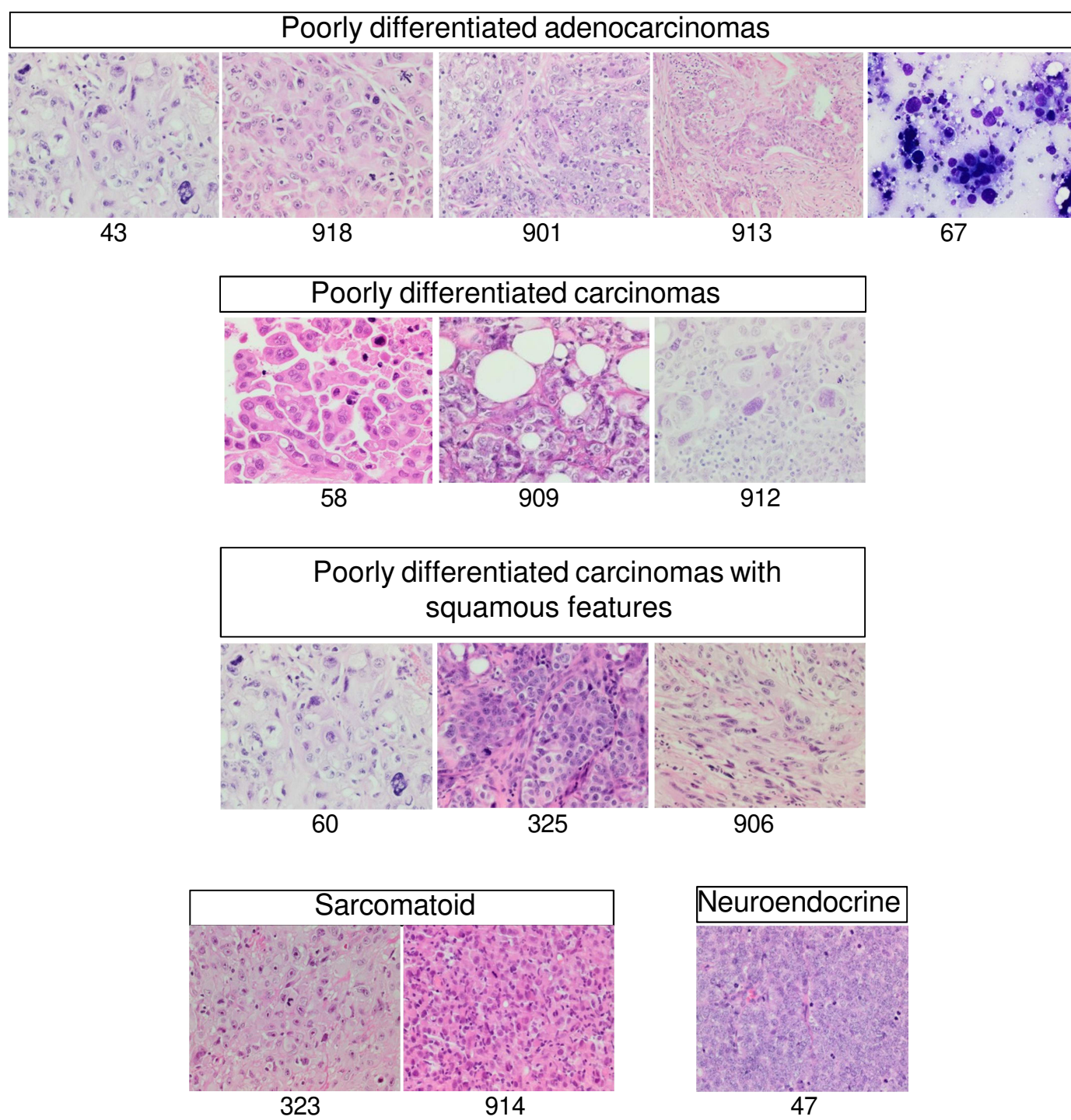
***Brundu et al.***

*Mutation of the axon guidance gene PLXNB2 sustains proliferative autonomy and confers invasive properties to stem cells isolated from Cancers of Unknown Primary*

## APPENDIX

### Table of Contents

- **Appendix Figure S1.** Histological images of CUP samples analyzed in this study  
*Page 2*
- **Appendix Figure S2.** Supplemental data on cell collapsing assays with PlxnB2 mutants  
*Page 3*
- **Appendix Figure S3.** Ligand binding assays comparing wild-type and G842C-PlxnB2  
*Page 4*
- **Appendix Figure S4.** Supplemental images of cell migration/invasion assays  
*Page 5*

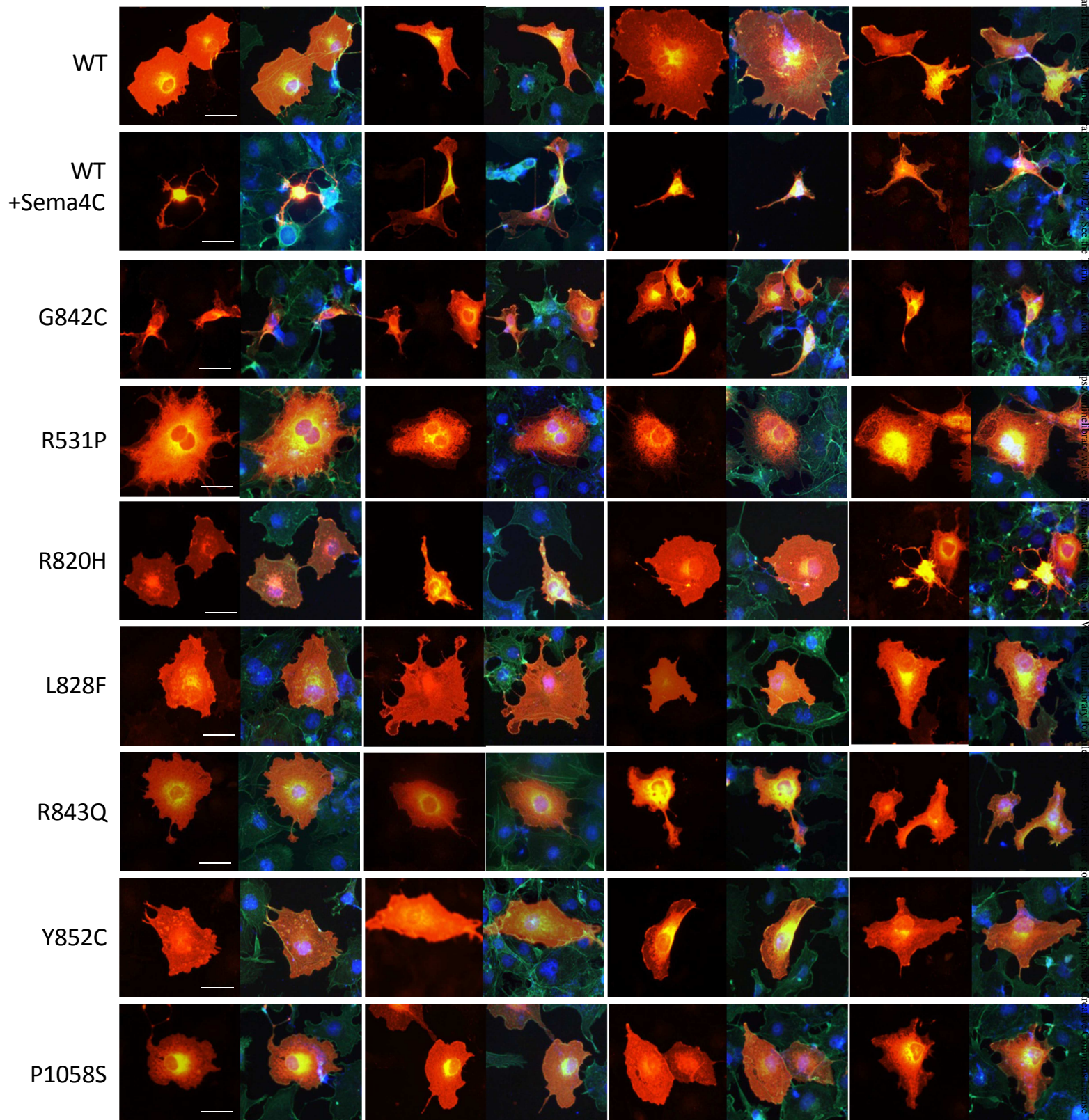
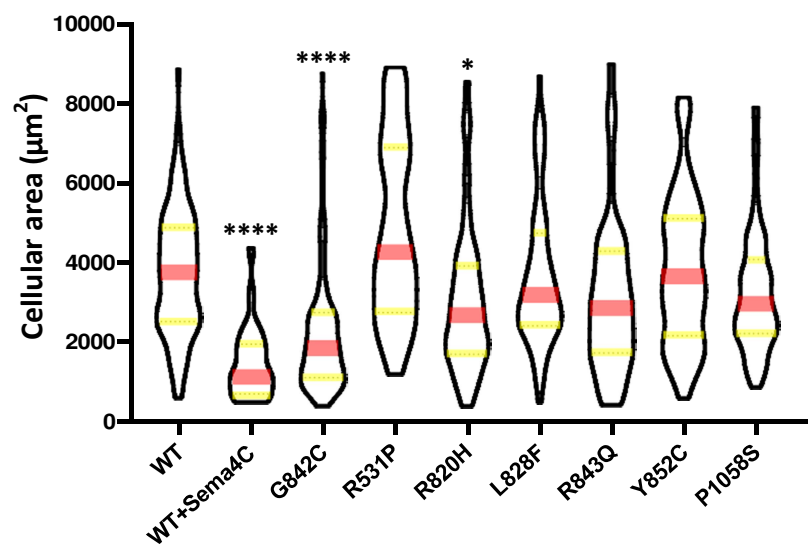


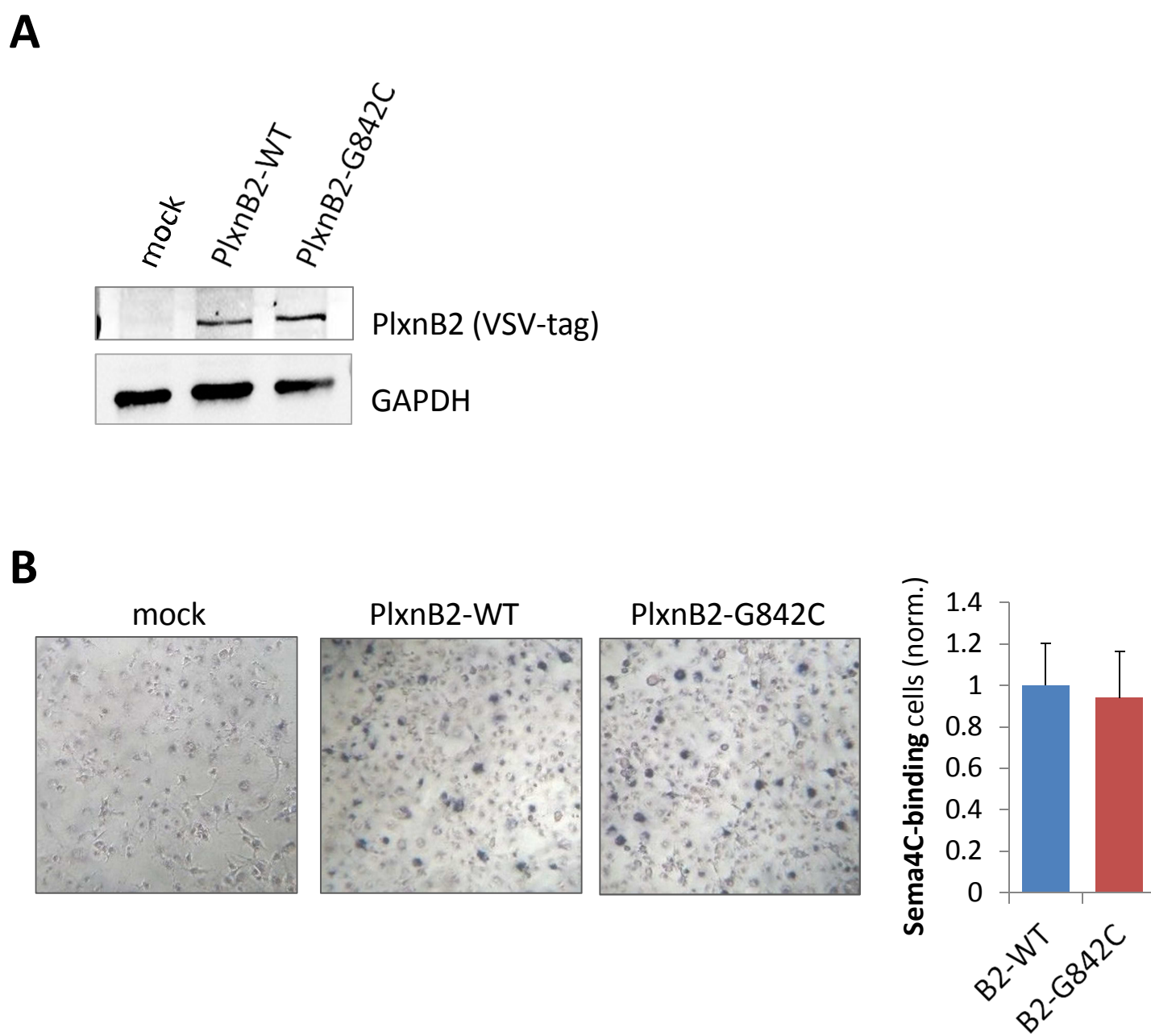
**Appendix Fig. S1. Histological images of CUP samples analyzed in this study.** Histological images (40x) of sectioned paraffin-embedded CUP samples (e.g. AGN43, AGN918, etc., as indicated in Table EV1), stained by Hematoxylin-Eosin staining, except for panel N (AGN67) which is a cytological specimen stained by Diff-Quick.



**Appendix Fig. S2. Collapsing assays with PlxnB2 mutants.** Immunofluorescence analysis of transfected COS7 cells expressing wild-type or mutant PlxnB2 variants. Cells expressing the WT receptor were also stimulated with 1  $\mu\text{g/ml}$  Sema4C, in order to provide an internal positive control of the typical collapsing response. The area of around 50-100 representative cells per each condition (from multiple independent microscopic fields and experiments) was measured by ImageJ software, and the individual values were included in a violin plot (shown on the right), where the red line indicates the median value and yellow lines mark the first and third quartiles of the population. The statistical analysis was done by one-way ANOVA, comparing each of the groups with that of unstimulated WT cells: \*\*\*\* $p < 0.0001$ ; \* $p < 0.05$ .

Below is shown a wide series of representative fields for the diverse conditions, presented as pairs of one image displaying anti-PlxnB2 staining in red, and one corresponding merge of fluorescent channels (including green staining of F-actin by phalloidin, and nuclear DAPI staining in blue). Scale bar: 50 $\mu\text{m}$ .



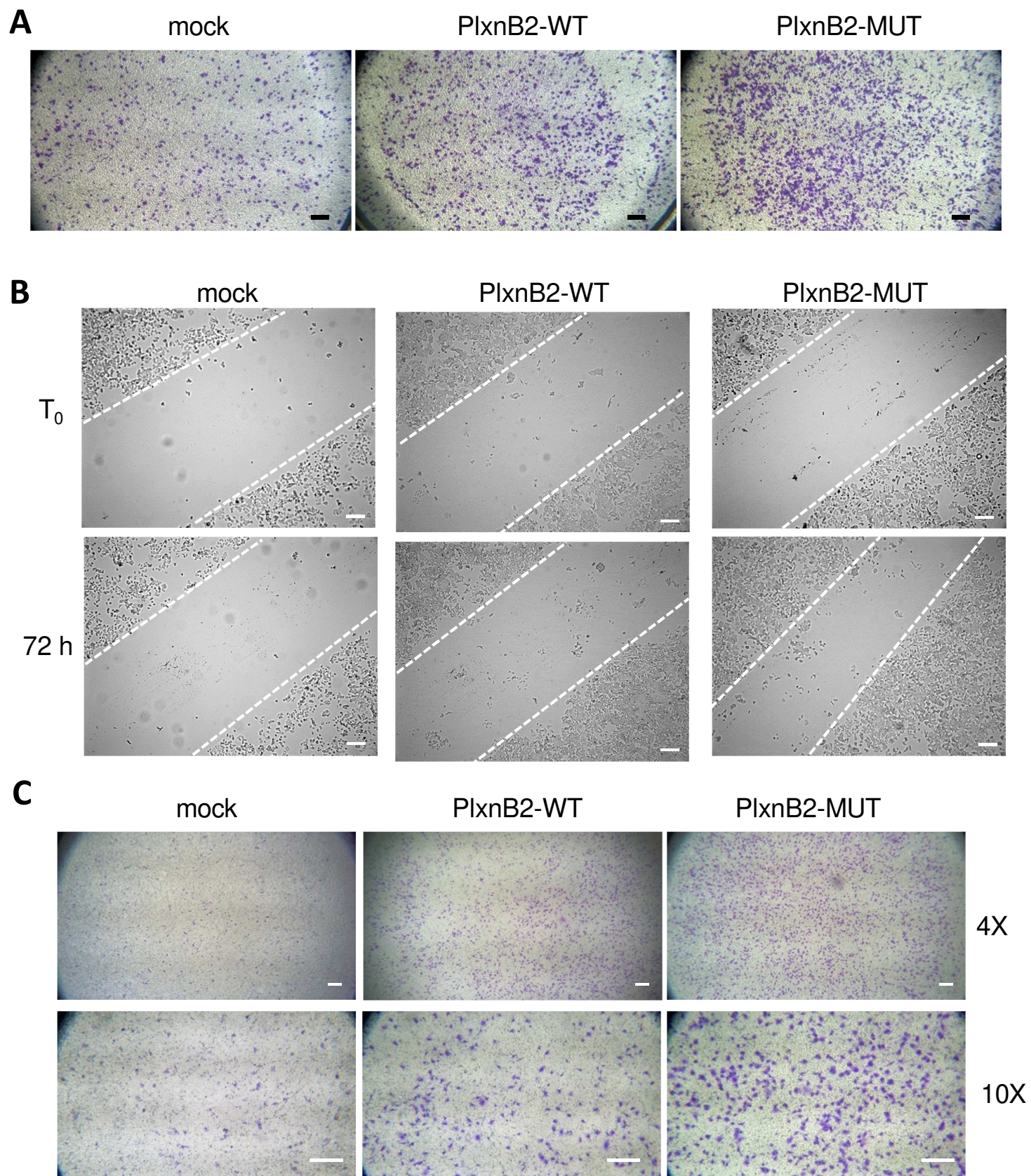


**Appendix Fig. S3. Ligand binding assays comparing wild-type and G842C-PlxnB2**

(A) The comparable expression of VSV-tagged wild-type (WT) and mutated PlxnB2 in transfected COS7 cells was verified by western blotting.

(B) Representative images of COS7 cells transfected with WT or G842C-mutated PlxnB2 (analyzed in the previous panel) and probed with alkaline phosphatase-conjugated Sema4C, its cognate ligand, revealing comparable receptor binding. The graph at the bottom shows normalized mean values  $\pm$  SD of the AP-labeled cell quantified in three experiments (after subtracting mock background signal).





**Appendix Fig. S4. Supplemental images of cell migration and cell invasion assays.**

(A) Representative microscopic images of MCF-7 cells, either mock transfected, or overexpressing WT or G842C-mutated PlxnB2, migrated through Transwell inserts, then fixed and stained with crystal violet. Scale bar: 200  $\mu$ m. Data quantification and statistical analysis across multiple replicates is shown in main Fig. 7B. (B) Representative phase contrast microscopic images of monolayers of the same MCF-7 cells as above (either mock, or overexpressing WT or G842C-mutated PlxnB2) immediately after scratching at T<sub>0</sub>, and then after 72 hours wound closure. Scale bar: 200  $\mu$ m. Data quantification and statistical analysis across multiple replicates is shown in main Fig. 7C. (C) Representative microscopic images (at 4x or 10x magnification) of AS906 cells (either mock, or overexpressing WT or G842C-mutated PlxnB2), which had invaded across Matrigel-coated Transwell inserts, and were then fixed and stained with crystal violet. Scale bars: 200  $\mu$ m. Data quantification and statistical analysis across multiple replicates is shown in main Fig. 7E.