EBioMedicine 46 (2019) 133–149

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Research paper

CTC phenotyping for a preoperative assessment of tumor metastasis and overall survival of pancreatic ductal adenocarcinoma patients

Yukun Sun [a], Guangdong Wu [b,c], Kok Suen Cheng [a], Anqi Chen [a], Kuang Hong Neoh [a], Shuiyu Chen [a], Zhewen Tang [a], Poh Foong Lee [d], Menghua Dai [c,**], Ray P.S. Han [a,e,*]

[a] College of Engineering, Peking University, Beijing 100871, China
[b] Dept of Hepatopancreatobiliary Surgery, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Institute for Precision Medicine, Tsinghua University, Beijing 102218, China
[c] Dept. of General Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China
[d] Dept. of Mechanical & Materials Engineering, University Tunku Abdul Rahman, Bandar Sungai Long, Selangor, Malaysia
[e] Integrated Chinese & Western Medicine Oncology Research Center, Jiangxi University of Traditional Chinese Medicine, Nanchang, Jiangxi 330004, China.

ARTICLE INFO

Article history:
Received 17 March 2019
Received in revised form 27 June 2019
Accepted 16 July 2019
Available online 30 July 2019

Keywords:
Pancreatic cancer
CTC phenotyping blood test
PDAC metastatic assessment
PDAC overall survival assessment

ABSTRACT

Background: The evaluation for surgical resectability of pancreatic ductal adenocarcinoma (PDAC) patients is not only imaging-based but highly subjective. An objective method is urgently needed. We report on the clinical value of a phenotypic circulating tumor cell (CTC)-based blood test for a preoperative prognostic assessment of tumor metastasis and overall survival (OS) of PDAC patients.
Methods: Venous blood samples from 46 pathologically confirmed PDAC patients were collected prospectively before surgery and immunoassayed using a specially designed TU-chip™. Captured CTCs were differentiated into epithelial (*E*), mesenchymal and hybrid (*H*) phenotypes. A further 45 non-neoplastic healthy donors provided blood for cell line validation study and CTC false positive quantification.
Findings: A validated multivariable model consisting of disjunctively combined CTC phenotypes: "*H*-CTC≥15.0 CTCs/2ml OR *E*-CTC≥11.0 CTCs/2ml" generated an optimal prediction of metastasis with a sensitivity of 1.000 (95% CI 0.889–1.000) and specificity of 0.886 (95% CI 0.765–0.972). The adjusted Kaplan-Meier median OS constructed using Cox proportional-hazard models and stratified for *E*-CTC < 11.0 CTCs/2 ml was 16.5 months and for *E*-CTC ≥ 11.0 CTCs/2 ml was 5.5 months (HR = 0.050, 95% CI 0.004–0.578, $P = .016$). These OS results were consistent with the outcome of the metastatic analysis.
Interpretation: Our work suggested that *H*-CTC is a better predictor of metastasis and *E*-CTC is a significant independent predictor of OS. The CTC phenotyping model has the potential to be developed into a reliable and accurate blood test for metastatic and OS assessments of PDAC patients.
Fund: National Natural Science Foundation of China; Zhejiang Province Science and Technology Program; China Scholarship Council.

Abbreviations: AIC, Akaike information criterion; AJCC, American Joint Committee on Cancer; AUC, Area under the curve; BSA, Bovine serum albumin; CE, Capture efficiency; CT, Computed tomography; CTC, Circulating tumor cell; CU, Capture unit; C-index, Concordance; DAPI, 4′,6-diamidino-2-phenylindole; DMEM, Dulbecco's Modified Eagle Medium (DMEM); E-cad, E-cadherin; EDTA, Ethylenediaminetetraacetic acid; EMT, Epithelial-mesenchymal transition; *E*, *M*, *H*, *T*, Epithelial, Mesenchymal, Hybrid, Total; FBS, Fetal bovine serum; HR, hazard ratio; IMDM, Iscove's Modified Dulbecco's Medium; LOOCV, Leave-one-out cross-validation; mAUC, Multivariable area under the curve; MET, Mesenchymal-epithelial transition; mLBR, Multivariable binary logistic regression; mROC, multivariable receiver operating characteristic; NPV, Negative predictive value; OS, Overall survival; PBS, Phosphate buffered saline; PCR, Polymerase chain reaction; PDAC, Pancreatic ductal adenocarcinoma; PDMS, Polydimethylsiloxane; PFA, Paraformaldehyde; PI, Propidium iodide; PPV, Positive predictive value; RFS, Relapse free survival; ROC, Receiver operating characteristic; SD, Standard deviation; STR, Short tandem repeat; TMN, Tumor-Node-Metastasis; TU, Triangular units; VIF, Variance inflation factor; WBC, White blood cell.

* Corresponding author at: Integrated Chinese & Western Medicine Oncology Research Center, Jiangxi University of Traditional Chinese Medicine, Nanchang, Jiangxi 330004, China.
** Corresponding author.
E-mail addresses: daimh@pumch.cn (M. Dai), ray-han@jxutcm.edu.cn (R.P.S. Han).

https://doi.org/10.1016/j.ebiom.2019.07.044

Research in context

Evidence before this study

Pancreatic ductal adenocarcinoma (PDAC), which comprises the majority of all pancreatic cancers is often diagnosed during late stages when metastasis has already occurred. With surgery as the only hope of treatment, a computed tomography (CT) scan for evaluating the suitability of surgical resectability is about 80% accurate, and often, adversely affected by subjective human interpretations. This motivates the need for a more accurate and objective appraisal of surgical resectability. Circulating tumor cells (CTCs) represent a promising biomarker for assessing PDAC tumor metastasis and survival prognosis, however, current research effort is focused on CTCs without considering the uniquely different roles played by each CTC phenotype in the pathogenesis of tumor metastasis.

Added value of this study

We report on a blood test that uses a microfluidic chip to harvest CTC phenotypes from peripheral blood to discriminate local and metastatic PDAC tumors. The metastatic cutoff limits we obtained possess superior sensitivity and specificity. Further, this objective tool can be used to perform a prognostic assessment of the overall survival (OS) of PDAC patients. The median survival of a metastatic patient with epithelial CTCs exceeding the cutoff limit is <6 months but for non-metastatic patients who subsequently underwent surgery, the survival prediction increases to almost 1.5 years.

Implications of all the available evidence

The use of the CTC phenotypic count instead of the popular total count leads to an enhanced accuracy in the preoperative prediction of tumor metastasis and also, the prognostic prediction of overall survival of PDAC patients. Therefore, a CTC phenotype-based blood test has the potential to be developed into an accurate and reliable tool for the assessment of local and metastatic PDAC tumors to complement the traditional imaging methods for tumor staging and surgical resectability and also, the OS prognosis of pancreatic cancer patients.

1. Introduction

Every year, a quarter million people worldwide will develop pancreatic ductal adenocarcinoma (PDAC), which comprises nearly 90% of all solid neoplasms of the pancreas and some 70% will die within a year of diagnosis [1,2]. Due to the late presentation and lack of effective therapies, PDAC in most patients manifests as a locally advanced or metastatic cancer with surgical resectability as the only hope of treatment [3]. Today, multidetector computed tomography (CT) scans can predict surgical resectability with an 80% accuracy [2,4], but the prediction is subjective and highly dependent on human interpretations, particularly, for situations involving occult and/or small-volume metastases that may not be visible during imaging but only to be found during a surgical exploration [5] and the consequent risk of a margin-positive (R1/R2) resection [6]. Clearly, an enhanced accuracy in assessing tumor staging and resectability is urgently needed to improve the prognosis and survival rates of pancreatic cancer patients.

One promising approach is to utilize circulating tumor cells (CTCs) as a biomarker to predict tumor metastasis and survival prognosis of patients. Current methods involve using the *total* CTC count to stratify local and metastatic tumors [7,8] and also, for monitoring relapse and survival [9–11]. Ignoring the phenotypic heterogeneity of the CTC population is contrary to the pathogenesis of cancer, which considers a dynamic differentiation of CTCs [12] between epithelial (E) and mesenchymal (M) states through a transition hybrid (H) state.

In an attempt to develop a reliable and accurate blood test for the metastatic and survival assessments of PDAC patients, we resorted to phenotypic profiling of CTCs to cogitate the uniquely different role played by each CTC phenotype in tumor metastases. We carried out a rigorous cell line validation study using 5 pancreatic and 2 non-pancreatic cell lines to characterize our specially designed microfluidic chip for capture efficiency, capture purity, CTC false positives and biomarker selection for CTC phenotyping. We employed a microfluidic platform to segregate and assay CTCs harvested from patients' blood into their 3 phenotypes; E, M and H with the total CTC given by $T = E + M + H$. The E and M phenotypes express prototypical markers, such as E-cadherin (E-cad) and vimentin, respectively [13], whereas, the H-phenotype consists of tumor cells undergoing the epithelial-mesenchymal transition (EMT) or mesenchymal-epithelial transition (MET) program and thus, expresses both markers simultaneously but still, regarded to be distinct from the E and M phenotypes [14,15]. For a preoperative assessment of tumor metastasis, we showed that our CTC phenotyping count is superior to that of using the total CTC count. The CTC blood test we have developed can be used to complement traditional imaging methods to further enhance the accuracy and reliability of PDAC tumor staging and resectability assessments. Additionally, we have developed another CTC phenotyping tool that can be used for an assessment of the overall survival (OS) and relapse free survival (RFS) prognostic predictions of PDAC patients.

2. Materials and methods

2.1. TU-chip™ design and system setup for harvesting CTCs

For a fast and effective capture of CTCs in a peripheral blood sample, a microfluidic chip consisting of several thousand micron-sized *triangular units* (TU) was used. The chip, aptly named as the TU-chip™ was designed using the AutoCAD software (Autodesk Inc., San Rafael, CA) and fabricated via a soft lithography process with a substrate thickness of 25 μm at CapitalBio Corp (Beijing, China). A 10:1 weight-ratio mixture of polydimethylsiloxane (PDMS, Sylgard 184, Dow Corning, USA) prepolymer with a curing agent was degassed, poured into the mold and cured at 60 °C for 4 h. The PDMS layer was peeled out, punched with access holes and bonded to a microscope glass slide via an oxygen plasma treatment. The micropillars inside the PDMS chip were examined for flaws using a scanning electron microscopy (SEM, Hitachi S-4800). The microfluidic system setup [16] consisted of the chip, tubing, connectors, reservoirs, syringes and syringe pumps (Longer Pump, Baoding, Hebei, China). The flow process can be viewed and captured in realtime using an inverted microscope (Leica Microsystems, DM IL LED). Prior to starting an experiment, the TU-chip™, all tubing, connectors and syringes were primed by flushing with phosphate buffered saline (PBS) (Wisent Corporation, Cat# 311–010-CL), together with 8 mM ethylenediaminetetraacetic acid (EDTA) and 1% bovine serum albumin (BSA) (Wisent Corporation, Cat# 800095-QG) to eliminate contaminants and air-bubbles inside the system.

2.2. Cell culture and size measurement

To facilitate the design of the capture chamber that includes the placement of triangular micropillars in the TU-chip™, we used 7 cancer cell lines sourced from the Cell Resource Center, Peking Union Medical College (head-office for the National Infrastructure of Cell Line Resource): 5 pancreatic cell lines; 3 from primary tumors (BxPC-1, MIAPaCa-2, Panc-1) and 2 from metastatic tumors (CFPAC-1 from liver metastasis and AsPC-1 from ascites), and 2 non-pancreatic cell lines; human lung alveolar adenocarcinoma (A549) and breast

adenocarcinoma (MDA-MB-231). The cell lines were checked for mycoplasma contamination by polymerase chain reaction (PCR) and cell culture, and their species origins confirmed by PCR. The identity of a cell line was authenticated via a short tandem repeat (STR) profiling (FBI, CODIS). The AsPC-1 cell line was maintained with RPMI 1640 (Wisent Corporation, Cat# 350–005-CL), the CFPAC-1 cell line with Iscove's Modified Dulbecco's Medium (IMDM) (Gibco, Cat# 12440053), the BxPC-3, MIAPaCa-2, Panc-1 and MDA-MB-231 cell lines with Dulbecco's Modified Eagle Medium (DMEM) (Wisent Corporation, Cat# 350–319-020-CL), and the A549 cell line with McCoy's 5A (Wisent Corporation, Cat# 317–011-CL) at 37 °C and 5% CO_2. All culture media were supplemented with 10% fetal bovine serum (FBS) (Wisent Corporation, Cat# 086–150-CL) and 1% penicillin-streptomycin (Wisent Corporation, Cat# 450–201-EL). The cultured cells were harvested by treating with 0.25% trypsin-EDTA (Wisent Corporation, Cat# 325–043-el) and their diametral measurements collected. Cell suspension was diluted with approximately 500 cells in 200 μl and then put into one well of a 96-well plate. Pictures of cells were taken by a CCD camera (Leica DFC450) on the microscope (Leica Microsystems, DM IL LED) and the cell size analyzed with ImageJ software (RRID: SCR_003070, https://imagej.nih.gov/ij/).

2.3. Finite element flow simulations

Finite element simulations of the fluid flow inside the capture chamber of the TU-chip™ were carried out to study the cell flow pathway predictions and the integrity of the captured cells. The simulation was performed using COMSOL Multiphysics software (COMSOL Inc., Stockholm, Sweden, RRID: SCR_014767) for the analysis of fluid flows, cell streamline patterns, flow velocities and shear rates. A laminar flow model was used because of the micron-size dimensions and a maximum flow rate of 1 ml/h at the inlet with an open boundary at the outlet.

2.4. Cell line validation study

The A549 cancer cell line was used to characterize the capture efficiency (CE) of the TU-chip™ under varying flow rates and cell concentrations. The cancer cells were stained with CellTracker Red CMTPX Dye (Invitrogen, Cat# C34552) to distinguish them from other blood cells prior to spiking into blood samples sourced from healthy volunteers. Blood samples were first centrifuged at 700 *g* for 5 min with the plasma discarded and diluted with buffer solution and then, spiked with stained cancer cells in concentrations of approximately 300 cells per 2 ml of blood and gently premixed before the experiment. After PBS washing, cell counting was performed using a hemocytometer or if the numbers were manageable, the counting would be manually carried out under a microscope. The capture efficiency (CE) of the chip is defined as the ratio of the total number of captured cells to the total number of input cells. To find an optimal flow rate for the capture, we tested the CE under five different flow rates: 0.25 ml/h, 0.5 ml/h, 1 ml/h, 1.5 ml/h and 2 ml/h using the A549 cell line with five replicates. Similarly, using the A549 cell line, the CE of the chip under varying cell concentrations was determined using 50, 100, 150 and 200 cells per 2 ml perfused through the chip at a flow rate of 1 ml/h. Lastly, to evaluate the CE of different cell lines, the number of cells for AsPC-1, BxPC-3, CFPAC-1, MIAPaCa-2, Panc-1, A549 and MDA-MB-231 was set at 300 per 2 ml under a flow rate of 1 ml/h. Both the experiments for the different cell concentrations and cell lines were performed in triplicate.

For the selection of marker for the identification of CTC phenotypes, the five pancreatic cancer cell lines were first cultured in 96-well plate until 60–70% confluency. The cultured cells were then washed with PBS twice, and then, fixed by a 4% paraformaldehyde (PFA) flow for 15 min followed by another three PBS washes. The fixed cells were permeabilized by streaming 0.1% Triton X-100 in PBS for 10 min followed by three PBS washes. The permeabilized cells were incubated with BlockAid Blocking Solution (Life Technologies, Cat# B10710) for 30 min to minimize nonspecific bindings. The cells were then perfused with antibodies: 4′,6-diamidino-2-phenylindole (DAPI) (Thermo Fisher Scientific Cat# D1306, RRID: AB_2629482), PE conjugated anti-EpCAM (Thermo Fisher Scientific Cat# MA1–10198, RRID: AB_11154356) or PE conjugated anti-pan-CK (Abcam Cat# ab52460, RRID: AB_870750), or Alexa Fluor 488 conjugated anti-CD45 (Thermo Fisher Scientific Cat# MHCD4520, RRID: AB_10392555), Alexa Fluor 555 conjugated anti-E-cadherin (Abcam Cat# ab206878, RRID: AB_2801591) and Alexa Fluor 647 conjugated anti-vimentin (Abcam Cat# ab195878, RRID: AB_2801592). All antibodies were diluted as per the manufacturer's instruction. The cells were then incubated at 4 °C overnight in the dark. Before imaging, the stained cells were washed with PBS. Pictures of the cells were taken by a CCD camera (Leica DFC450) on the microscope (Leica Microsystems, DM IL LED) with appropriate fluorescent filter cubes. A comparison of the varying levels of EpCAM expressions from high to low of the 5 pancreatic cancer cell lines was done under the same staining conditions and imaged with the same objective, excitation light intensity and exposure time. The fluorescence intensity was measured using ImageJ software (RRID: SCR_003070, https://imagej.nih.gov/ij/).

2.5. Cancer cell recovery and cell viability assessment

To recover tumor cells inside the capture chamber, the chip was first flushed with Pluronic® F127 (Sigma-Aldrich, Cat# P2443-250G) to avoid nonspecific adhesion of cells, followed by a reversed PBS flush to collect the captured cells in a 96-well plate connected to the recovery outlet. The recovery efficiency of the chip was defined as the ratio of the total number of recovered cells to the total number of captured cells and the experiments was performed using A549 cells at varying concentrations of 50, 100, 200, 500 and 1000 cells per 2 ml. For each concentration, the experiment was performed in triplicate.

The viability of the captured cells was assessed using a Cellstain double staining kit (Sigma-Aldrich, Cat# 05411) for 2 situations: unrecovered captured cells and recovered captured cells, and compared with cells before capture. Briefly, the cells were incubated at room temperature for 30 min with 0.15 μM of calcein-AM and 0.8 μM of propidium iodide (PI) simultaneously. Viable cells were stained with calcein-AM that emitted green fluorescence whereas dead cells were stained with PI that emitted red fluorescence. The cell viability was quantified as the ratio of the total number of viable cells to the total number of cells, which were both counted under a fluorescence microscope. The experiments were performed in triplicate. To further confirm the viability of the cell, the recovered cells were cultured with McCoy's 5A medium (Wisent Corporation, Cat# 317–011-CL) at 37 °C and 5% CO_2.

2.6. Participants, sample size and sample collection

This observational study was carried out from September 2015 to September 2016 at Peking Union Medical College Hospital, Beijing, China. Forty-six pathologically confirmed, treatment-naive PDAC patients divided into 2 cohorts: 35 local (stage I-III) and 11 metastatic (stage IV) cases were recruited. The mean age of the two groups were 59.9 ± 8.6 and 58.5 ± 5.0 years, respectively, and the percentage of males were 57.1% and 45.5%, respectively. A further 45 non-neoplastic healthy donors provided blood: 20 as negative controls for quantifying the CTC false positive issue and the remaining 25 for carrying out chip characterization spiking experiments. All subjects were anonymously coded with informed consent obtained and the work was performed in accordance to the Declaration of Helsinki with the research protocol approved by the hospital institutional review board.

The proposed CTC blood test predicts the onset of metastasis in the 46 PDAC patients and therefore, the control group refers to the 35 local patients and the case group pertains to the 11 metastatic patients. Selecting the desired margin of error for the specificity and sensitivity of

the respective control and case cohorts, we can accurately determine the required sample size of the local and metastatic groups [17]. In our work, we prescribed a margin of error of 0.10 with specificity and sensitivity of 0.90 and 0.99 respectively, to reflect the desire for a good accuracy, especially, in minimizing false negatives to achieve a highly sensitive test. The computed minimum sample size turned out to be 39 (35 local and 4 metastatic patients). Further, adopting an effect size of 1.01 [7,8,18] to secure a statistical power of 0.80 with $\alpha = 0.05$ and using a 3:1 control-to-case ratio for a greater precision [19], the computed minimum sample size was 44 (33 local and 11 metastatic patients). Clearly, our sample size of 46 (35 local and 11 metastatic patients) met these requirements and also, satisfied the "rule of thumb" sample size for a pilot study [20].

All blood samples were collected in vacutainer tubes containing the anticoagulant EDTA. Only 4 ml of blood was drawn from each patient given that late stage cancer patients are often unable to provide 7.5 ml of blood [21]. The first 2 ml of the extracted blood sample was discarded in order to minimize the contamination of endothelial cells from the venipuncture. Blood was shipped or stored at room temperature but was processed within 4 h of collection. The pathology of resected tumors was reviewed by a trained pathologist for an analysis of the tumor size, grade, nodal status, perineural and perivascular invasion, carcinoma cell embolus, and staging based on the 8th edition of the American Joint Committee on Cancer (AJCC) Tumor–Node–Metastasis (TMN) classification [22].

2.7. Immunofluorescent assaying of circulating tumor cells

From each patient, 2 ml of blood was processed using our TU-chip™ at room temperature. Blood samples were centrifuged at 700*g* for 5 min

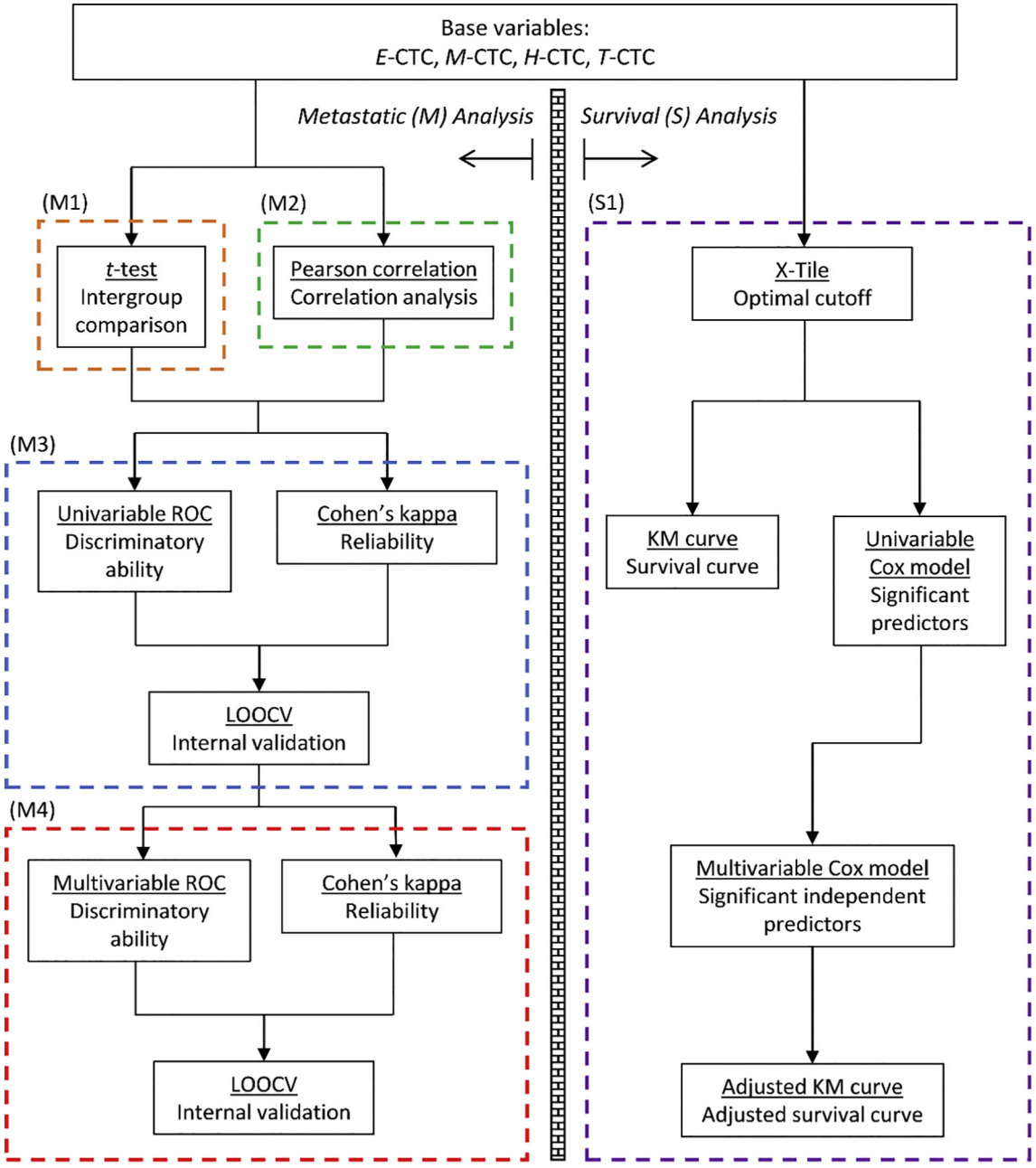

Fig. 1. Statistical flowchart for the 2 CTC phenotyping models: metastatic (M) and survival (S) analyses. (M1) Intergroup *t*-test comparisons for the CTC phenotype count of local and metastatic PDAC patients. (M2) Correlation analysis between each CTC phenotype with other known PDAC biomarkers. (M3) Univariable and (M4) multivariable analyses for CTC phenotype discrimination of local and metastatic patients. (S1) Survival analysis for OS and RFS.

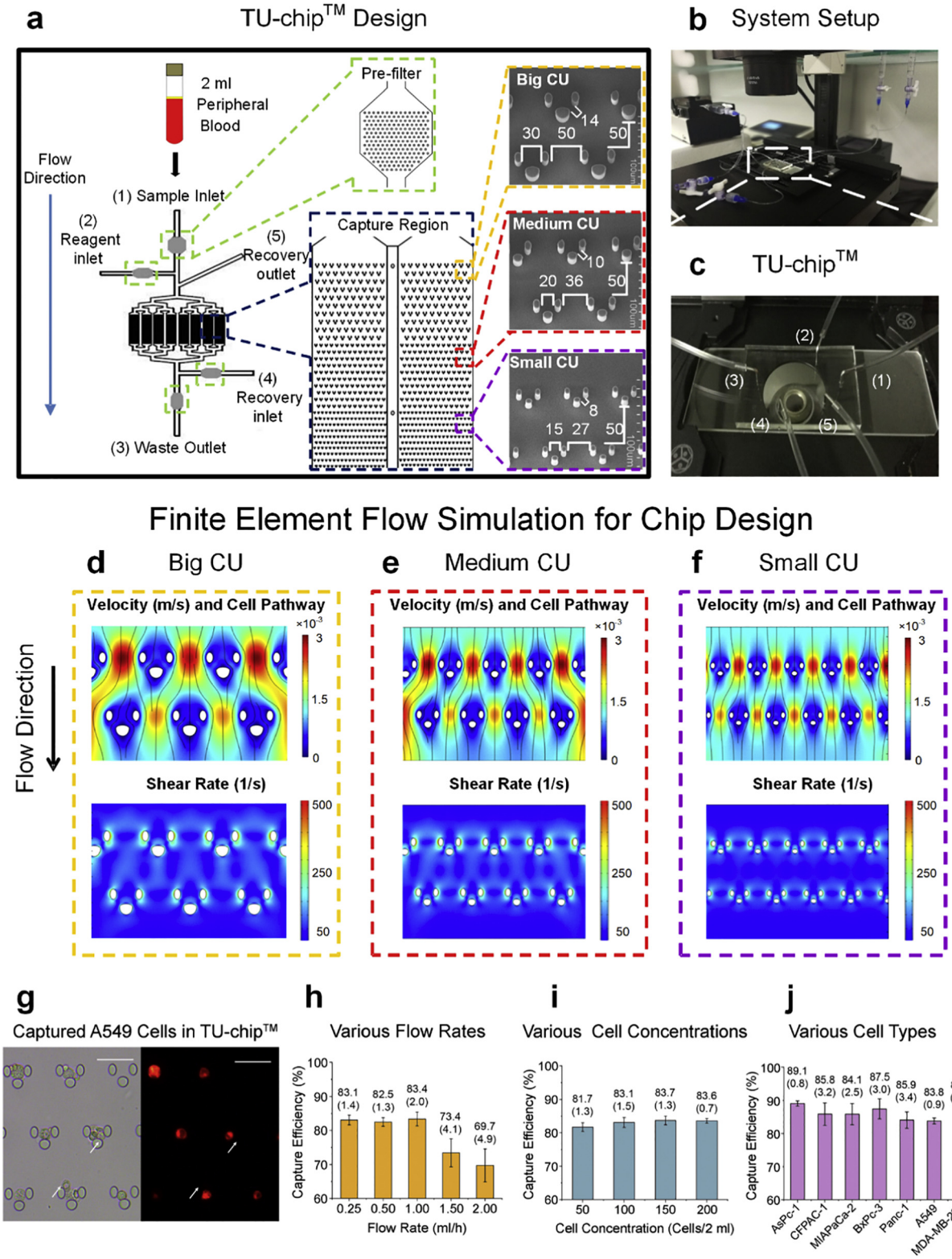

Fig. 2. Microfluidic system setup and design of TU-chip™. (a) TU-chip™ consists of 5 entry-exit points (sample inlet (1), reagent inlet (2), waste outlet (3), recovery inlet (4) and recovery outlet (5)), a pre-filter to minimize clogging from large particles and a capture region that consists of 8 chambers × 693 triangular capture units/chamber arranged in 3 "big-medium-small" placement configurations (SEM image). (b) System setup for a rapid harvesting and identification of CTCs. (c) Zoom-in view of the TU-chip™. (d-f) Finite element simulation of fluid flows inside the TU-chip™ using the COMSOL Multiphysics software (COMSOL Inc., Stockholm, Sweden): velocity (m/s) and cell pathway (top) and shear rate (1/s) (bottom) of the (d) big, (e) medium and (f) small placements of the capture units. (G-J) Spiked blood experiments for chip characterization. (g) Capture of A549 cells marked red and contaminated blood cells (indicated by white arrows), scale bar: 40 μm. (h-j) Capture efficiency (mean (SD)) of TU-chip™ for varying flow rates (5 repeats) (h), cell concentrations (3 repeats) (i) and cell types (3 repeats) (j). Note: the error bars represent the standard deviation.

Table 1
Measured size data of cancer cell lines for the design of the TU-chip™ capture chamber.

Cell line	A549	MDA-MB-231	AsPC-1	BxPC-3	CFPAC-1	MIAPaCa-2	Panc-1
Min (μm)	10.0	10.0	6.0	6.0	7.0	6.0	8.0
Max (μm)	19.0	22.0	21.0	23.0	29.0	24.0	28.0
Mean (μm)	14.3	14.7	14.5	12.3	16.7	13.0	18.6
SD (μm)	1.7	2.1	3.6	2.9	3.8	2.5	3.6

with the supernatant serum discarded and the pellet diluted with a buffer of volume ratio 1:1. The sample was syringe-pumped into the microfluidic chip at a rate of 1 ml/h. After captured, the CTCs were identified by an immunofluorescent staining technique. The isolated cells inside the chip were first washed with PBS and then, fixed by a 4%-paraformaldehyde (PFA) flow for 15 min followed by another PBS-

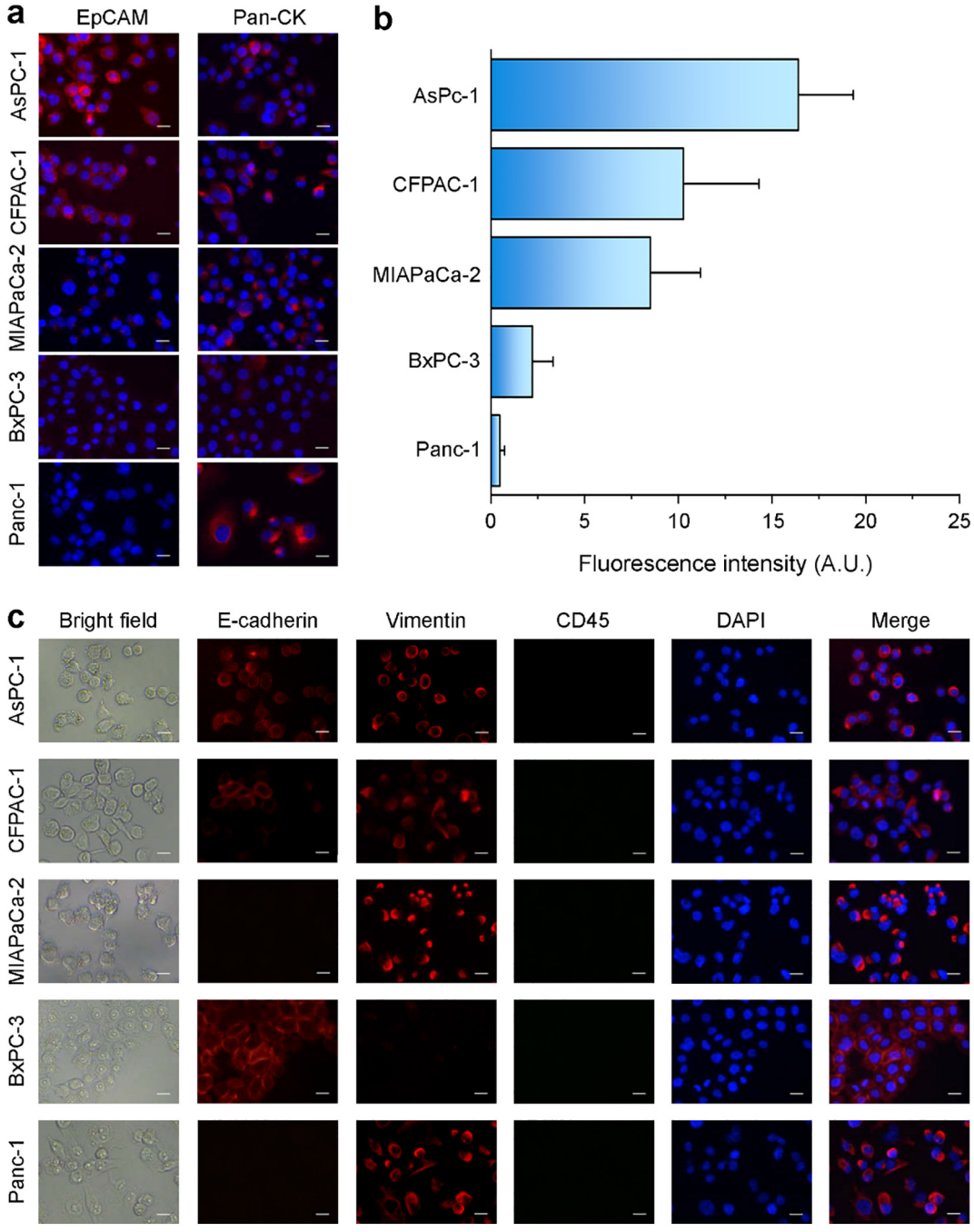

Fig. 3. Cell-line validation study. (a) Immunofluorescence staining of 5 pancreatic cancer cell lines: AsPC-1, CFPAC-1, MIAPaCa-2, BxPC-3 and Panc-1 with EpCAM and pan-CK. (b) Quantitative comparison of the fluorescence intensity of EpCAM for the 5 pancreatic cancer cell lines. (c) Immunofluorescence staining of the 5 cell lines with *E*-cad, vimentin, CD45 and DAPI showing the varying epithelial/mesenchymal expressions of the cell lines. Scale bar: 20 μm. Note: A.U. denotes arbitrary unit. The error bars represent the standard deviation.

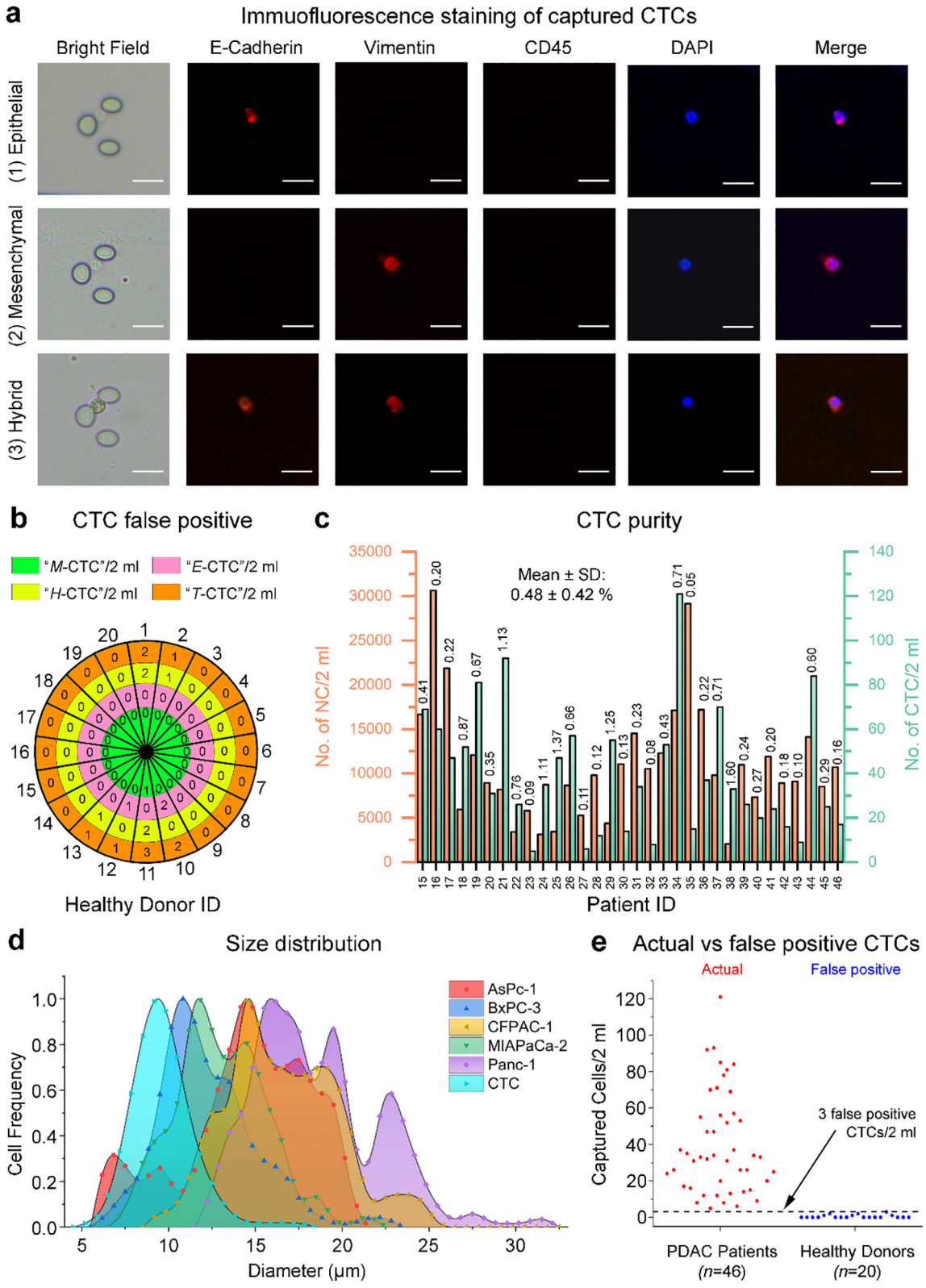

Fig. 4. CTC characterization: phenotype identification, false positive, capture purity, cell-size distribution and actual CTC count versus false positive CTC count. (a) Immunofluorescence staining of captured CTCs: (1) epithelial CTCs (DAPI+/CD45−/E-cad+/vimentin-), (2) mesenchymal CTCs (DAPI+/CD45−/E-cad−/vimentin+), (3) hybrid CTCs (DAPI+/CD45−/E-cad+/vimentin+). Scale bar: 20 μm. (b) False positive CTC counts from healthy donors ($n = 20$). The outer most ring of numbers refers to the healthy donor ID and the inner band of rings indicate the no. of false positive CTC phenotypes. A maximum of 3 false positive CTCs/2 ml was obtained. (c) CTC capture purity as measured by the number of captured WBCs from PDAC patients ($n = 32$). The figure above each individual bar denotes the CTC/WBC percentage ratio for a patient. (d) CTC size distribution: captured CTCs versus the 5 pancreatic cancer cell lines (AsPc-1, BxPC-3, CFPAC-1, MIAPaCa-2 and Panc-1). (e) Comparing actual CTC count from PDAC patients ($n = 46$) with false positive CTC count from healthy donors ($n = 20$) for a proper handling of the false positive issue. The dashed line represents the detection noise level at 3 false positive CTCs/2 ml of blood.

wash for 10 min. The fixed cells were permeabilized by streaming 0.1% Triton X-100 in PBS for 15 min, followed by a final PBS-wash for 10 min. The permeabilized cells were incubated with BlockAid Blocking Solution (Life Technologies, Cat# B10710) for 30 min to minimize non-specific bindings. The chip was then perfused with DAPI (Thermo Fisher Scientific Cat# D1306, RRID: AB-2629482) to stain the cell nuclei, Alexa Fluor 488 conjugated anti-CD45 (Thermo Fisher Scientific Cat# MHCD4520, RRID: AB_10392555) to mark white blood cells, EMT markers Alexa Fluor 555 conjugated anti-*E*-cad (Abcam Cat# ab206878, RRID:AB_2801591) to mark epithelial cells and Alexa Fluor 647 conjugated anti-vimentin (Abcam Cat# ab195878, RRID:AB_2801592) to mark mesenchymal cells [23,24]. All antibodies were diluted in accordance to the manufacturer's instruction and flowed at 500 μl/h into the chip, which was then incubated at 4 °C overnight in the dark. Before imaging, the stained cells were washed with PBS. The TU-chip™ was imaged using an inverted fluorescence microscope under a 20× magnification in both the bright field and fluorescence modes with appropriate fluorescent filter cubes. Morphologically, CTCs were defined as having a circular or oval shape, combined with immunofluorescence staining results of DAPI+/CD45−/E-cad+/vimentin- for epithelial CTC, DAPI+/CD45−/E-cad−/vimentin+ for mesenchymal CTC or DAPI+/CD45−/E-cad+/vimentin+ for hybrid CTC. No size restriction was imposed in our identification of CTCs. To eliminate bias, all cells were enumerated by 2 double-blind analysts.

2.8. Statistical modeling, validation and survival analysis

The data was analyzed using SPSS version 20.0 (SPSS, Inc., Chicago, IL, USA, RRID: SCR_002865), MATLAB version 2017a (The MathWorks, Inc., MA, USA, RRID: SCR_001622) and the R software version 3.5.2 (R Project for Statistical Computing, RRID: SCR_001905, http://www.r-project.org/). For all continuous variables, the Kolmogorov-Smirnov test was used to determine the normality of the variables. If a variable was normally distributed, it was presented as a "mean (standard deviation; SD)"; otherwise, as a "median (interquartile range; IQR)". The size distribution of the 7 cell lines were compared using the Fisher's exact test of homogeneity [25] and their CE were compared using a one-way ANOVA. For intergroup comparisons between local and PDAC patients, the Pearson chi-square test, independent samples *t*-test and Mann-Whitney-*U* test were employed respectively, for the categorical, normally distributed, and non-normally distributed variables. Spearman rank correlation and Pearson correlation were used to test the bivariate correlation of non-parametric and parametric data, respectively. For the univariable CTC count model, a receiver operating characteristic (ROC) curve to differentiate between local and metastatic patients was drawn-up and an optimum cut-off value was obtained based on the highest Youden's index. The area under the curve (AUC) was computed to assess the overall discriminatory ability. To investigate the combined diagnostic capabilities of the CTC phenotype combinations, two multivariable models were formulated: multivariable binary logistic regression (mBLR) and multivariable receiver operating characteristic curve (mROC) [26]. For the former model, various combinations of two CTC phenotypes were used as predictors in the logistic equation. For mROC, it was setup using the logical 'OR', 'AND' operators for all possible disjunctive and conjunctive combinations of two phenotypes for a total of 592 combinations listed below.

- $E \geq 10.5$ CTCs/2 ml OR $M = 50$ combinations
- $E \geq 10.5$ CTCs/2 ml OR $H = 48$ combinations
- $M \geq 21.0$ CTCs/2 ml OR $E = 50$ combinations
- $M \geq 21.0$ CTCs/2 ml OR $H = 48$ combinations
- $H \geq 15.0$ CTCs/2 ml OR $E = 50$ combinations
- $H \geq 15.0$ CTCs/2 ml OR $M = 50$ combinations
- $E \geq 10.5$ CTCs/2 ml AND $M = 50$ combinations
- $E \geq 10.5$ CTCs/2 ml AND $H = 48$ combinations
- $M \geq 21.0$ CTCs/2 ml AND $E = 50$ combinations
- $M \geq 21.0$ CTCs/2 ml AND $H = 48$ combinations
- $H \geq 15.0$ CTCs/2 ml AND $E = 50$ combinations
- $H \geq 15.0$ CTCs/2 ml AND $M = 50$ combinations

The process involved holding one fixed with the cutoff based on an individual ROC, and the other varying with the cutoff changing in unit increments. The optimum cutoff for varying CTC phenotypes in the multivariable model was determined based on the highest Youden's index. The multivariable area under the curve (mAUC) was computed using MATLAB by first scaling the rectangle covered by the bottom-left and upper-right points of the mROC and its interior to a 1×1 square to enable the area under the scaled curve to be evaluated (Fig. S1). For both of the models, the validity measures (Youden's index, sensitivity, specificity, positive/negative predictive value (PPV, NPV), and accuracy) were obtained and the reliability of the CTC phenotyping combinations was tested against the gold standard of pathology through the Cohen's kappa coefficient.

The leave-one-out cross-validation (LOOCV) technique was used to validate both univariable and multivariable models. Although being more computationally intensive, the LOOCV technique yields greater accuracy in validating a model's performance compared to the traditional holdout validation technique [27]. Further, LOOCV uses the available data in a more efficient way and is not limited by the need to arbitrarily separate them in to training and test datasets. The 95% confidence interval was computed using 1000 bootstrap iterations.

For the survival analysis of overall survival (OS) and relapse-free survival (RFS), an optimal cutoff point that maximizes the difference (χ^2 value) between survival curves for each CTC phenotype was first determined using X-Tile version 3.6.1 [28]. The survival curves were compared using the log-rank test. A univariable Cox regression model was used to determine possible significant predictors of OS and RFS and those with $P < .05$ were then entered into a multivariable Cox regression model to identify independent significant predictors. A maximum of four predictors was entered into the multivariable Cox regression

Table 2
CTC capture platforms and number of captured CTCs for pancreatic cancer.

Reference	Capture Platform	Capture Method	Captured CTCs	
			Reported Units	Normalized Units[a]
Wei et al. [1]	CytoQuest™ CR system + microfluidic chip	Vimentin or EpCAM immobilized microfluidic chip	0–23 vimetin+ CTCs/4 ml	0–5.75 vimentin+ CTCs/ml
Ankeny et al. [2]	NanoVelcro Chip	EpCAM coated nanosubstrate	0–48 CTCs/4 ml	0–12 CTCs/ml
Poruk et al. [3]	ISET	Size-based filtration using 8 um pores	1–251 *E*-CTCs/ml 1–16 *M*-like CTCs/ml	1–251 *E*-CTCs/ml 1–16 *M*-like CTCs/ml
Gemenetzis et al. [7]	ISET	Size-based filtration using 8 um pores	0–25 *T*-CTCs/ml	0–25 *T*-CTCs/ml
Our work	TU-chip™	Microfluidic size-based capture with segregated spacings	5–121 *T*-CTCs/2 ml	2.5–60.5 *T*-CTCs/ml

[a] Normalized to per ml for the purpose of comparison. ISET = Isolation by Size of Epithelial Tumor Cell; CTC = circulating tumor cell; *E* = epithelial; *M* = mesenchymal; *H* = hybrid; *T* = total = $E + M + H$.

model in order to ensure the validity of the model by satisfying the minimum of 5 events per variable rule of thumb [29]. Adjusted Kaplan-Meier survival curves stratified for the *E*-CTC were constructed using the multivariable Cox proportional-hazard model and the method for calculating adjusted survival using the mean of the TMN stage, *M*-CTC and *H*-CTC. A two-sided $P < .05$ was considered as statistically significant. A flowchart is provided (Fig. 1) to better understand and follow the logic of our statistical methodology.

3. Results

3.1. TU-chip™ design characterization and performance assessment

The TU-chip™ has 5 inlet-outlet points: 3 inlets (sample inlet, reagent inlet, and recovery inlet) and 2 outlets (waste outlet and recovery outlet), a pre-filter to minimize clogging from large particles and a capture region consisting of 8 chambers × 693 capture units (CU)/chamber = 5544 CUs, each consisting of a group of 3 elliptical micropillars placed in a triangular configuration (Fig. 2a). The chip was connected to the rest of the system via tubing and connectors with the blood flow driven by syringe pumps (Fig. 2b–c). To trap a wide range of tumor cells, 3 types of CU characterized as small, medium and big that correspond to gap sizes of 8–30 μm were used (Fig. 2a). The overall chip design was based on size measurement data of 5 pancreatic cell lines (AsPC-1, BxPC-3, CFPAC-1, MIAPaCa-2 and Panc-1) and 2 non-pancreatic cell lines (A549 and MDA-MB-23) and the results are listed in Table 1. The size data of a representative tumor cell ranges from 6.0 (min) to 29.0 (max) with 14.9 (mean) μm and these figures are consistent with reported sizes of CTCs [30,31]. Based on these measurements, the largest and smallest distances between any 2 micropillars in the 3 types of CU varied from 15 to 30 and 8–14 μm, respectively. Further, our TU-chip™ allows red blood cells (4–6 μm) to pass through the capture chamber relatively unimpeded.

Fig. 2d–f depicts the finite element flow simulation inside the capture chamber to facilitate the assessment of streamline patterns (to predict cell flow pathways) and shear rates (to ascertain the viability of captured cells). The velocity and cell pathway simulation showed that for all 3 CUs, the flow velocity increases between adjacent CUs but decreases inside the interior area of a CU. When the cell pathway intersected a CU, it indicates a possible capture of a CTC within that particular CU (Fig. 2d–f, top sub-panels). Therefore, the placements of CUs were spatially gradated in compliant with the results of the FEM streamline distributions. The computed shear rate inside a CU ranges from 0 to 400 1/s (Fig. 2d–f, bottom sub-panels) and this falls within the range of normal human physiological states (from 10 1/s in veins to 2000 1/s in the smallest arteries) [32] to give us the confidence that the captured CTCs are able to largely remain intact and viable (Fig. S2).

Cell line validation studies involving spiked blood experiments were carried out using 5 pancreatic and 2 non-pancreatic cancer cell lines (Fig. 2g) to assess the performance of TU-chip™ for the CE under varying flowrates (five replicates), cell concentrations and cell types (both triplicate) (Fig. 2h–j). The results showed a relatively stable CE at about 83% as the flowrate increases from 0.25–1.00 ml/h but quickly dropped to 69.7% as the flowrate rises to 2.00 ml/h. The reason for the drop in the CE as the flowrate increases above the threshold is simply because more cells were escaping, either as whole (intact) cells or cell fragments caused by the higher shear rates of the faster flows (Fig. S3). The optimal flow rate was determined to be 1.00 ml/h with a peak CE of 83.4% (2.0%) (Fig. 2h). The CE was found to be relatively stable at about 83% for varying cell concentrations of 50–200 cells/2 ml (Fig. 2i). This result suggests that the chip should be able to capture CTCs consistently for both local and advanced stage PDAC patients with an increasing CTC count as the disease progresses. Observe in Fig. 2j that the CE of TU-chip™ stayed consistently high at above 82% for the 5 pancreatic cancer cell lines and 2 non-pancreatic cancer cell lines. This is despite testing on pancreatic cell lines with varying EpCAM expressions that range from high (AsPC-1) to low (BxPC-1 and Panc-1) (Fig. 3a) and the fluorescence intensity of EpCAM (Fig. 3b). Unlike an affinity-based platform that possesses a biomarker dependent capture performance, our capture method is wholly size-based and therefore, its CE stayed unaffected by varying levels of the EpCAM expressions. Further, our testing showed that the CEs for the 7 cell lines were statistically indifferent ($P = .332$, one-way ANOVA) even though they were measured using cell lines with a significant difference in their size distribution ($P < .001$, Fisher's exact test of homogeneity).

Table 3
Patient characteristics and CTC counts of local and metastatic PDAC.

Patient variable	Local	Metastatic	*P* value
No. (%)	35 (76)	11 (34)	—
Age, mean (SD), *y*	59.9 (8.6)	58.5 (5.0)	0.604
Males, No. (%)	20 (57)	5 (46)	0.497
Albumin, mean (SD), *g*	42.7 (3.3)	41.0 (3.0)	0.130
CA19-9 serum, median (IQR), *U/ml*	106 (35—273)	355 (95—634)	0.116
CEA serum, mean (SD), *ng/ml*	5.1 (4.0) (*n* = 27)	3.7 (3.9) (*n* = 7)	0.407
CA242 serum, mean (SD), *U/ml*	42.9 (45) (*n* = 25)	79.3 (60.6) (*n* = 10)	0.059
Location of tumor: head/(body or tail), No. (%)	24 (69)/11 (31)	5 (46)/6 (54)	0.641
Size of tumor, mean (SD), *cm*	3.6 (1.6)	—	—
Differentiation grade, No. (%)			
Well	4 (12)	—	—
Moderate	19 (54)	—	—
Poor	12 (34)	—	—
Not specified	—	11 (100)	—
Regional lymphnode metastasis, No. (%)			
N0	18 (51)	—	—
N1	17 (49)	—	—
Nx	—	11 (100)	—
Perineural invasion, No. (%)			
Yes	23 (66)	—	—
No	12 (34)	—	—
Not specified	—	11 (100)	—
Perivascular invasion, No. (%)			
Yes	3 (9)	—	—
No	32 (91)	—	—
Not specified	—	11 (100)	—
Carcinoma cell embolus, No. (%)			
Yes	8 (23)	—	—
No	27 (77)	—	—
Not specified	—	11 (100)	—
TNM Stage, No. (%)			
I	1 (3)	—	—
IIA	16 (46)	—	—
IIB	17 (48)	—	—
III	1 (3)	—	—
IV	—	11 (100)	—
Surgery, No. (%)			
Whipple	23 (66)	—	—
Distal pancreatectomy	11 (31)	—	—
Palliative surgery	—	4 (36)	—
Others	1 (3)	3 (28)	—
No Surgery	—	4 (36)	
Resection Margin - R0, No. (%)			
Yes	19 (54)		—
No	16 (46)		—
Not Specified		—	—
Chemotherapy, No. (%)			
Yes	30 (86)	3 (27)	< 0.001
No	5 (14)	6 (55)	
Not specified	0 (0)	2 (18)	
Number of CTCs/2 ml			
Total CTCs/2 ml, mean (SD)	29.7 (20.3)	74.1 (22.5)	< 0.001
Total CTCs: minimum, maximum	5, 85	47, 121	—
Epithelial CTCs/2 ml, mean (SD)	5.4 (3.9)	20.6 (12.8)	0.003
Mesenchymal CTCs/2 ml, mean (SD)	15.3 (10.9)	29.9 (10.5)	< 0.001
Hybrid CTCs/2 ml, mean (SD)	8.7 (7.0)	23.6 (9.7)	< 0.001

This result indicates that the CE of TU-chip™ is independent of the cell size.

3.2. Marker selection for CTC phenotyping, CTC false positive and capture purity issues

It is important to select appropriate markers for the immunofluorescent (IF) identification of CTC phenotypes. The epithelial marker, pan-cytokeratin (pan-CK) is widely used for the identification of CTCs without any reference to their phenotypes. Since our work involves CTC phenotypes, we argue that E-cad represents a better choice as evident from the following test. We stained our 5 pancreatic cancer cell lines with DAPI, CD45, vimentin and 2 epithelial markers, pan-CK and E-cad and the results are shown in Fig. 3a, c. All 5 cancer cell lines stained positive for pan-CK, but only 3 cancer cell lines, AsPC-1, BxPC-3 and CFPAC-1 did so for E-cad. As for the mesenchymal marker vimentin, only BxPC-3 cells were negatively stained for vimentin. These staining results are consistent with published reports [33–35] describing either the epithelial or mesenchymal properties of these cell lines. For example, MIAPaCa-2 and Panc-1 cell lines possess mainly mesenchymal characteristics [36] and this was reflected in the staining results of *E*-cad−/vimentin+. However, the use of pan-CK as the epithelial marker would falsely these 2 cell lines as having a hybrid phenotype. Hence, our staining antibody panel consisted of the nucleus marker DAPI, the leukocyte marker CD45, the epithelial marker E-cad and the mesenchymal marker vimentin.

Having selected appropriate markers for the identification of CTC phenotypes, we proceeded to define our CTC phenotypes based on a combined morphologic criterion (round or oval) and the IF staining result. Also, due to the obvious size overlap between CTCs and white blood cells (WBCs) [37], we did not invoke a size cutoff in our CTC definition. The 3 CTC phenotypes were IF identified as follows (Fig. 4a): epithelial or *E*-CTC (DAPI+/CD45−/E-cad+/vimentin-), mesenchymal or *M*-CTC (DAPI+/CD45−/E-cad−/vimentin+) and hybrid or *H*-CTC (DAPI+/CD45−/E-cad+/vimentin+) with the total or *T*-CTC given by $T = E + M + H$.

WBCs can be similar in size to CTCs and often, inadvertently captured in large quantities, particularly when size-based methods such as our TU-chip™ are employed. Their presence pose 2 challenging problems: CTC false positive and capture purity and they require an accurate quantification so that they can be properly handled. The first issue of CTC false positives, in which WBCs are unintentionally labelled as CTCs (as per the immunostaining results) is complicated as it depends on whether the blood is drawn from healthy donors or pathologically confirmed cancer patients. We assessed false positive CTCs in our chip by using blood samples sourced from 20 healthy donors and apply the same sample processing and immunostaining steps (i.e. DAPI/CD45/E-cad/vimentin) as with a confirmed PDAC patient. To quantify as a false positive, we counted the number of captured false positive *E*-CTCs (DAPI+/CD45−/E-cad+/vimentin-), false positive *M*-CTCs (DAPI+/CD45−/E-cad−/vimentin+), false positive *H*-CTCs (DAPI+/CD45−/E-cad+/vimentin+). False positive CTCs were found in 6 of the healthy donors' blood with a maximum count of 3 false positive CTCs/2 ml of blood and hence, the false positive detection rate for the TU-chip™ is 30% (6/20) (Fig. 4b). As long as the cutoff limits of our statistical model are above the background noise of 3 false positive CTCs/2 ml, our CTC blood test results will not be affected.

The second issue is concerned with the CTC capture purity, which is defined as follows: [38].

$$\text{CTC Capture Purity} = \frac{(\text{Total CTC})_{\text{captured}}}{\text{Total } (\text{CTC} + \text{WBC})_{\text{captured}}},$$

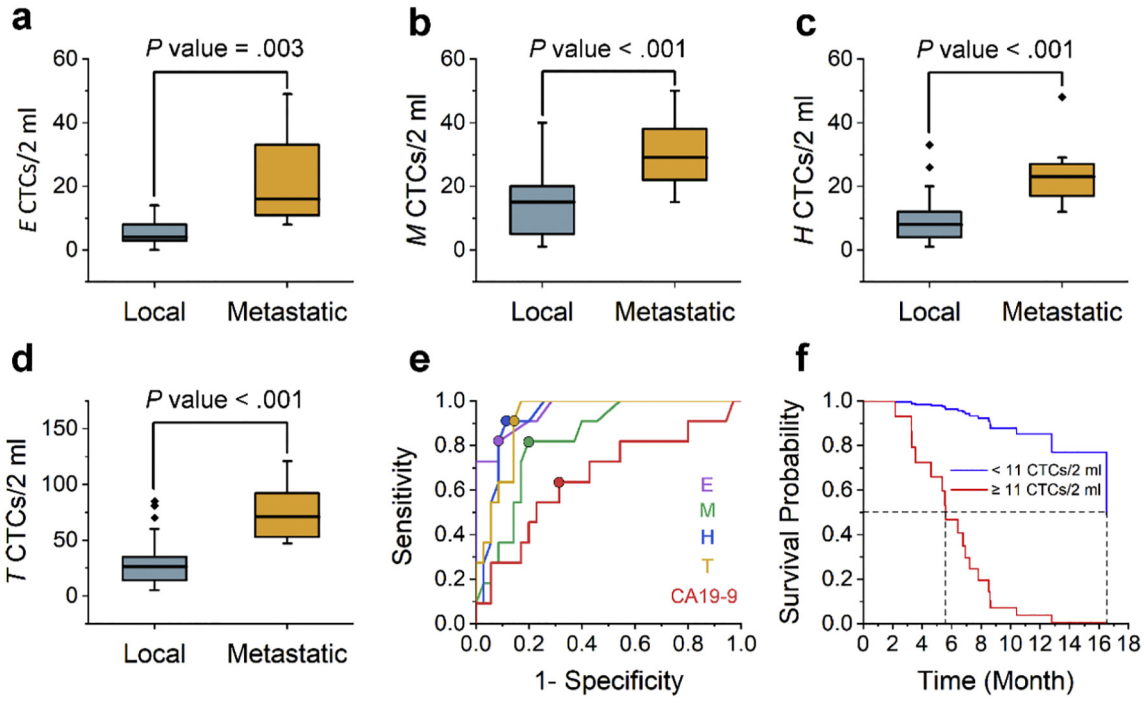

Fig. 5. Correlation of CTC phenotype counts with PDAC metastasis and overall survival. (a-d) Correlation between local and metastatic PDAC patients for (a) *E*-CTC, (b) *M*-CTC, (c) *H*-CTC and (d) *T*-CTC. *P* values shown are for *t*-test comparison between local and metastatic patients. The numerical values for the mean (SD) of each variable are listed in Table 3. (e) Comparison of ROC curves for *E*-CTC (cutoff = 10.5 CTCs/2 ml, AUC = 0.956 (95% CI 0.898–1.000; $P < .001$, Mann-Whitney *U* test)), *M*-CTC (cutoff = 21.0 CTCs/ml, AUC = 0.831 (95% CI 0.709–0.953; $P < .001$, Mann-Whitney U test)), *H*-CTC (cutoff = 15.0 CTCs/2 ml, AUC = 0.929 (95% CI 0.854–1.000; $P < .001$, Mann-Whitney U test)), *T*-CTC (cutoff = 42.0 CTCs/2 ml, AUC = 0.926 (95% CI 0.853–0.999; $P < .001$, Mann-Whitney U test)), and CA19–9 (cutoff = 184.6 U/ml, AUC = 0.658 (95% CI 0.462–0.855; $P = .116$, Mann-Whitney U test)). (f) Adjusted Kaplan-Meier OS stratified with respect to the *E*-CTC cutoff value of 11 CTCs/2 ml for PDAC patients ($n = 46$). The curves were adjusted using the mean of the TMN stage, *M*-CTC and *H*-CTC. Note: the box represents the first, second and third quartiles. The lower and upper ends of the whiskers represent the lowest and highest datum, respectively, but still within 1.5 of the IQR from the lower and upper quartile. The diamonds denote the outliers.

The issue arises because WBCs (CD45+ or CD45+/vimentin+) [39] were also undesirably captured in large quantities with our size-based approach and thus, it is necessary to quantify them for a proper handling. We evaluated the capture purity of 32 patients and obtained the mean (SD) capture purity of 0.48 (0.42)% (Fig. 4c). Despite the low capture purity of our chip, our size-based method still harvests significantly more CTCs than affinity-based enrichment platforms (Table 2). The low capture purity of our chip simply implies that we are also capturing disproportionately more WBCs. This does not cause a problem for our CTC identification because the gradated segregation trapping of the cells ensures that the view of our capture chamber is clean and uncluttered with no build-up of debris. This makes the identification of CTCs and their phenotypes through their IF expressions easy and accurate.

The filtration-based ISET technology reported significantly higher false positive CTCs that are an order of magnitude higher compared to our result [40]. This is hardly surprising considering the large pile of cluttered cells in membrane filtration technologies for cell capture.

3.3. CTC phenotype characterization and correlation to metastasis

The baseline characteristics of the 46 PDAC patients comprising of 35 local (stage I-III) and 11 metastatic (stage IV) cases are listed in Table 3 (more details in Table S1). CTCs were found in 100% of the patients with a total CTC count ranging from 5 to 121 CTCs/2 ml of blood (comprising 5–85 CTCs/2 ml for local patients and 47–121 CTCs/2 ml for metastatic patients) (Table 3). The captured CTCs had a mean (SD) diameter of 9.6 (1.7) μm and their size distribution overlapped with those of the 5 pancreatic cell lines (Fig. 4d). No CTC clusters were observed. Also, a comparison of the false positive CTCs with the actual CTCs harvested from the 46 patients (Fig. 4e) showed that the outcome of the IF staining is significantly above the background noise for the former to adversely affect our CTC results.

Comparing the local and metastatic patients, it was clear that the latter have significantly higher CTC counts for all *E*,*M*,*H*,*T* categories ($P = .003$, *t*-test for *E* and $P < .001$, *t*-test for *M*,*H*,*T*) (Fig. 5a–d). However, no significant difference was observed in the most commonly used biomarker CA19–9 and other conventional biomarkers (CEA, CA242) and tumor location (Table 3). Additionally, there was no correlation of the conventional biomarkers (CA19–9, CEA, CA242) and tumor size with the CTC phenotypes (Fig. S4–7). These results suggest that CTC phenotype counts could potentially be a good independent marker for PDAC metastasis.

3.4. Univariable CTC phenotype count model

In this model, the CTC count variables for differentiating local and metastatic patients are *E*, *M*, *H* and *T* and their cross-validated optimal cutoff values, validity and reliability measures are listed in Table 4 (training data in Table S2). Among the 3 CTC phenotypes, the *H*-CTC possessed the highest sensitivity of 0.909 (95% CI 0.707–1.000) whereas, the *E*-CTC had the highest specificity of 0.914 (95% CI 0.811–1.000). When all the 3 CTC phenotypes are lumped together to form the *T*-CTC, the sensitivity rose to 1.000 (95% CI 1.000–1.000) but the specificity dropped to 0.829 (95% CI 0.691–0.943). Further, from an integrated assessment of discriminatory performance (Youden's index = 0.795) and reliability (kappa = 0.727 (95% CI 0.481–0.942), the *H*-CTC is ranked among the best of the 4 variables. The AUC for each CTC phenotype yielded a statistical prediction that ranges from "good-to-excellent": *E*-CTC: AUC = 0.956 (95% CI 0.898–1.000; $P < .001$, Mann-Whitney *U* test); *M*-CTC: AUC = 0.831 (95% CI 0.709–0.953; $P < .001$, Mann-Whitney U test); *H*-CTC: AUC = 0.929 (95% CI 0.854–1.000; $P < .001$, Mann-Whitney U test); *T*-CTC: AUC = 0.926 (95% CI 0.853–0.999; $P < .001$, Mann-Whitney U test). Compared with the performance of CA19–9, which the AUC was only 0.658 (95% CI, 0.462–0.855; $P = .116$, Mann-Whitney U test), the *H*-CTC could be a better biomarker for metastasis (Fig. 5e).

Table 4
Cross-validated cutoff values, validity and reliability for univariable and multivariable CTC phenotype count models.

CTC Phenotype/Combination	Univariable model				Multivariable model ("Fixed phenotype" ≥" Fixed threshold" OR "Varying phenotype")					
	E	*M*	*H*	*T*	*E* ≥ 10.5 CTCs/2 ml OR *M*	*E* ≥ 10.5 CTCs/2 ml OR *H*	*M* ≥ 21.0 CTCs/2 ml OR *E*	*M* ≥ 21.0 CTCs/2 ml OR *H*	*H* ≥ 15.0 CTCs/2 ml OR *E*	*H* ≥ 15.0 CTCs/2 ml OR *M*
Cutoff value[a] (CTCs/2 ml)	10.5	21.0	15.0	42.0	23.0	17.0	11.0	14.0	11.0	28.0
Youden's index	0.732	0.618	0.795	0.829	0.738	0.823	0.709	0.800	0.886	0.857
Sensitivity	0.818	0.818	0.909	1.000	0.909	0.909	0.909	1.000	1.000	1.000
	(0.539–1.000)	(0.556–1.000)	(0.707–1.000)	(1.000–1.000)	(0.710–1.000)	(0.692–1.000)	(0.727–1.000)	(1.000–1.000)	(0.889–1.000)	(1.000–1.000)
Specificity	0.914	0.800	0.886	0.829	0.829	0.914	0.800	0.800	0.886	0.857
	(0.811–1.000)	(0.662–0.923)	(0.771–0.974)	(0.691–0.943)	(0.706–0.944)	(0.809–1.000)	(0.653–0.941)	(0.667–0.921)	(0.765–0.972)	(0.714–0.971)
PPV	0.750	0.563	0.714	0.647	0.625	0.769	0.588	0.611	0.733	0.688
	(0.450–1.000)	(0.314–0.813)	(0.467–0.933)	(0.400–0.879)	(0.380–0.857)	(0.500–1.000)	(0.357–0.846)	(0.378–0.833)	(0.467–0.933)	(0.433–0.913)
NPV	0.941	0.933	0.969	1.000	0.967	0.970	0.966	1.000	1.000	1.000
	(0.843–1.000)	(0.839–1.000)	(0.897–1.000)	(1.000–1.000)	(0.889–1.000)	(0.882–1.000)	(0.893–1.000)	(1.000–1.000)	(0.966–1.000)	(1.000–1.000)
Accuracy	0.8913	0.804	0.891	0.870	0.848	0.913	0.826	0.848	0.913	0.891
	(0.783–0.957)	(0.696–0.913)	(0.783–0.978)	(0.783–0.957)	(0.739–0.935)	(0.826–0.978)	(0.717–0.924)	(0.739–0.935)	(0.826–0.978)	(0.804–0.978)
Kappa	0.710	0.535	0.727	0.698	0.638	0.775	0.597	0.535	0.788	0.742
	(0.438–0.901)	(0.252–0.803)	(0.481–0.942)	(0.469–0.898)	(0.403–0.864)	(0.529–0.945)	(0.335–0.825)	(0.251–0.804)	(0.556–0.950)	(0.494–0.928)
P value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

[a] In multivariable model, this cutoff value is the "Optimal cutoff value for the 'varying phenotype'".

3.5. Multivariable CTC phenotype count model

To further improve the performance metrics of the univariable model, we investigated all possible conjunctive and disjunctive combinations of CTC phenotypes using 2 multivariable models: mBLR and mROC. Correlation analysis showed that the *T*-CTC was highly correlated to the other 3 CTC phenotypes (Pearson $r > 0.75$) (Table S3) and the variance inflation factor (VIF) became large when included in the models (Table S4). Hence, to avoid the issue of multicollinearity [26,41], the *T*-CTC was dropped in favor of combinations involving only single CTC phenotypes of *E, M* and *H*.

The validated results of the mBLR analysis are depicted in Table S5 (training data in Table S6). The validated results of the mROC analysis for the "OR" operator are summarized in Table 4 (training data in Table S7), and the validated results for the "AND" operator are presented in Table S8 (training data in Table S9).

The changes of the validity measures (Youden's index, accuracy, sensitivity, specificity, PPV and NPV) with the varying phenotype cutoff in the mROC model for the 'OR' and 'AND' combination are shown in Figs. S8, S9, respectively. Focusing on the Youden's index, two patterns were observed corresponding to the 'OR' and 'AND' combinations.

For the 'OR' combination, the Youden's index initially increased with the cutoff until it attained a maximum value (Fig. S8). As the cutoff continued to further increase, the Youden's index either plateaued or decreased to a lower value before re-plateauing. On the other hand, the Youden's index for the 'AND' combination generally started off at a maximum value and then, decreased in a stepwise manner at higher cutoffs, i.e. with a plateau sandwiched between subsequent decrements (Fig. S9). For both the 'OR' and 'AND' combinations, the same best performance can sometimes be obtained by multiple cutoff points. For this situation, the first cutoff was chosen as the optimal for the varying phenotype since any subsequent cutoffs are a subset of the first cutoff (e.g. "≥ 15 CTCs/2 ml" includes "≥ 16 CTCs/2 ml", etc).

Between the "AND" and "OR" logical operators, the best diagnostic performance was achieved by the disjunctive ("OR") combination of '*H*-CTC≥15.0 CTCs/2 ml OR *E*-CTC≥11.0 CTCs/2 ml' with an improved sensitivity of 1.000 (95% CI 0.889–1.000), specificity of 0.886 (95% CI 0.765–0.972), PPV of 0.733 (95% CI 0.467–0.9333), NPV of 1.000 (95% CI 0.966–1.000), accuracy of 0.913 (95% CI: 0.826–0.978) and Cohen's kappa = 0.788 (95% CI 0.556–0.950). The diagnostic performance of the top 5 *optimized* combinations are presented in Table S10 – note that the optimal combinations represent the optimized varying phenotype cutoff for the 2 respective CTC phenotypic combinations, without taking into account the *non*-optimized cutoffs.

3.6. CTC phenotype count for OS and RFS analyses

We performed a prefatory study of the survival analysis on all 46 PDAC patients, who were followed for a maximum of 18 months after blood draw (Fig. S10). We carried out a survival analysis for OS for the

Table 5
Univariable and multivariable Cox regression analyses for overall survival during follow-up.

Patient variable	Overall survival					
	Univariable model		Multivariable model			
	HR (95% CI)	*P* value	Phenotype CTC count		Total CTC count	
			Concordance Index = 0.799 (0.045)[a] R^2 = 0.431, AIC = 110.72		Concordance Index = 0.754 (0.047) R^2 = 0.324, AIC = 114.62	
			HR (95% CI)	*P* value	HR (95% CI)	*P* value
Age, *y*	1.000 (0.947–1.055)	0.989				
Sex						
Female	Ref.					
Male	0.692 (0.278–1.724)	0.429				
Albumin, *g*	1.014 (0.875–1.176)	0.850				
CA19–9						
≥38 U/ml	Ref.					
<38 U/ml	0.800 (0.261–2.447)	0.695				
Location of Tumor						
Body or Tail	Ref.					
Head	1.077 (0.415–2.796)	0.879				
TNM Stage						
I	- (0-inf)	0.986	- (0-inf)	0.988	- (0-inf)	0.998
II[b]	Ref.		Ref.		Ref.	
III	6.257 (0.753–52.025)	0.090	11.263 (1.231–103.084)	0.032	9.044 (1.035–79.018)	0.046
IV	4.097 (1.534–10.943)	0.005	0.871 (0.194–3.910)	0.857	0.715 (0.175–2.920)	0.641
E-CTC						
≥11 CTCs/2 ml	Ref.		Ref.			
<11 CTCs/2 ml	0.109 (0.040–0.297)	<0.001	0.050 (0.004–0.578)	0.016		
M-CTC						
≥35 CTCs/2 ml	Ref.		Ref.			
<35 CTCs/2 ml	0.103 (0.038–0.284)	<0.001	0.321 (0.068–1.518)	0.152		
H-CTC						
≥16 CTCs/2 ml	Ref.		Ref.			
<16 CTCs/2 ml	0.222 (0.084–0.584)	0.002	3.561 (0.445–28.497)	0.231		
$E + M + H = T$-CTC[c]						
≥69 CTCs/2 ml	Ref.					
<69 CTCs/2 ml	0.116 (0.044–0.307)	<0.001				
$E + H = T_2$-CTC						
≥32 CTCs/2 ml	Ref.				Ref.	
<32 CTCs/2 ml	0.141 (0.053–0.373)	<0.001			0.096 (0.022–0.416)	0.002

OS, overall survival; HR, hazard ratio; CI, confidence interval; AIC, Akaike information criterion; Ref., reference; TNM, Tumor-Node-Metastasis; *E*-CTC, epithelial CTC; *M*-CTC, mesenchymal CTC; *H*-CTC, hybrid CTC; *T*-CTC = *E* + *M* + *H*-CTC; T_2-CTC = *E* + *H*-CTC; Variables which has a *P* value <.05 in univariable analysis was included in the multivariable analysis.

[a] $P = 0.131$ as compared to the concordance of the traditional approach.

[b] Stage II selected as the reference group since Stage I cancer has only 1 patient.

[c] Not included in the multivariable analysis due to multicollinearity.

46 patients and RFS for 35 local patients who underwent surgical resection. At the time of the last follow-up, 19 (41%) patients died with the overall median OS for all 46 patients was 12.8 months (95% CI 9.067–16.533). The univariable Cox regression analysis revealed the TNM stage, *E*-CTC, *M*-CTC, *H*-CTC and *T*-CTC count were significantly associated with OS. Following that, the multivariable analysis revealed that the *E*-CTC was a significant independent predictor for the OS with hazard ratio (HR) of 0.050 (95% CI 0.004–0.578) (Table 5). The adjusted Kaplan-Meier median for OS and stratified for *E*-CTC < 11 CTCs/2 ml was 16.53 months and for *E*-CTC ≥ 11 CTCs/2 ml 5.53 months (Fig. 5f). The univariable analysis showed the presence of carcinoma cell embolus and the *E*-CTC, *M*-CTC, *H*-CTC and *T*-CTC count were significantly associated with poorer RFS (Table S11). However, the multivariable analysis found none of the variables was a significant independent predictor of the RFS.

From unadjusted Kaplan-Meier curves stratified for all CTC phenotypes, we observed that some early stage patients with high CTCs (*E*-CTC > 11, *M*-CTC > 35, *H*-CTC > 16, *T*-CTC > 69 CTCs/2 ml) showed significantly poor OS (Fig. 6). Four early stage patients (Patient ID (TNM stage)) with high CTCs and by our assessment marked under the "poor OS" category (Patient ID 26 (IIA); 7 (IIB); 19 (IIB); 37(IIB)) (Table S1) died rapidly within 8.53, 5.53, 7.80 and 3.27 months (Fig. S10) respectively, after their tumor resection. One patient (Patient ID 16 (IIB)) considered to be a "medium-to-high" CTC case did not die after 12.63 months of follow-up, but developed a recurrence with 4 months of surgery. Of the 35 resected patients, 19 (54%) developed a recurrence and 11 (31%) died with an overall median RFS for all resected patients at 10.17 months (95% CI 7.671–12.669). From the Kaplan-Meier curves of RFS, we observed that patients with *E*-CTC > 8, *M*-CTC > 27, *H*-CTC > 13, *T*-CTC > 47 CTCs/2 ml possess significantly worse RFS (Fig. 7). There were a total of 12 patients who had higher CTCs in either one or more of their phenotypic or total CTCs than the cut-off limits possess worse RFS with 9 (75%) that developed a recurrence. Among them, 4 patients (Patient ID 7 (IIB); 19 (IIB); 26 (IIA); 37 (IIB)) had all 4 CTC phenotype counts beyond the cut-off limit and rapidly developed a relapse within 3.23, 1.17, 6.17, 1.70 months after their resection. Two patients (Patient ID 16 (IIB) and Patient ID 17 (IIB)) who had *M*, *H*, *T*-CTC beyond the cut-off developed a recurrence within 4 and 9.53 months, respectively after surgery. Three patient (Patient ID 3 (III); Patient ID 24 (IIA) and Patient ID 38 (IIB)) had 10 *E*-CTCs and developed a relapse within 4.90, 7.87 and 3.57 months of surgery.

4. Discussion

To develop an effective CTC-based blood test, we need to consider 3 major issues. The first issue is to select an appropriate device for

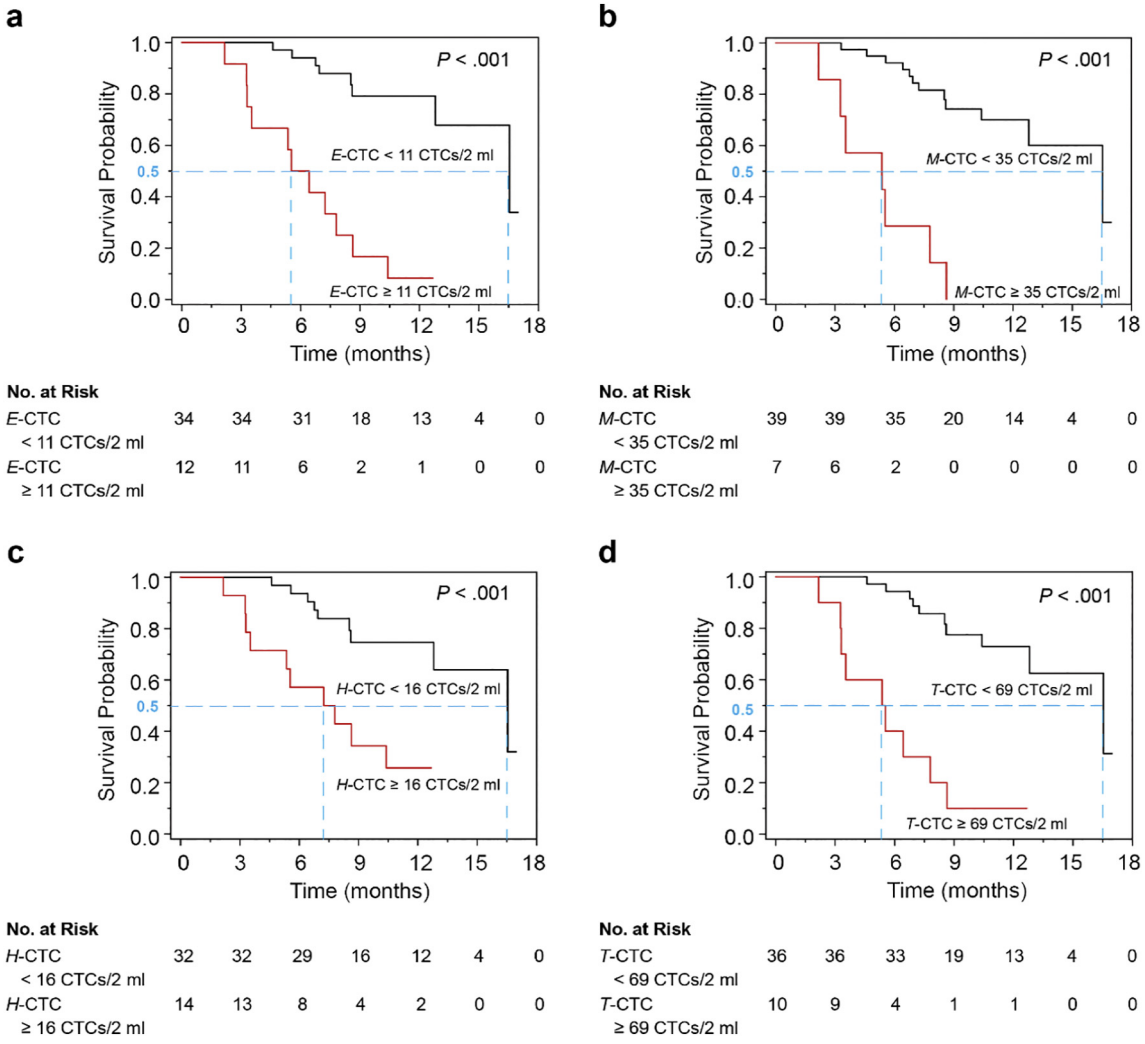

Fig. 6. Unadjusted Kaplan-Meier OS curves for PDAC patients ($n = 46$) and separated into 2 groups for (a) *E*-CTC, for patients with *E*-CTCs/2 ml < 11 vs ≥11, and the median OS was 16.53 months (95% CI 11.19–21.87) vs 5.53 months (95% CI 3.73–7.33). (b) *M*-CTC, for patients with *M*-CTCs/2 ml < 35 vs ≥35, the median OS was 16.53 months (95% CI 11.17–21.89) vs 5.37 months (95% CI 0.65–10.09). (c) *H*-CTC, for patients with *H*-CTCs/2 ml < 16 vs ≥ 16, the median OS was 16.53 months (95% CI 11.17–21.89) vs 7.32 months (95% CI 3.07–11.39). (d) *T*-CTC, for patients with *T*-CTCs/2 ml < 69 vs ≥69, the median OS was 16.53 months (95% CI 11.17–21.89) vs 5.37 months (95% CI 2.27–8.47). *P* values are for the log-rank test between survival curves.

harvesting CTCs and in our case, we employed our TU-chip™ to perform the CTC capture. The performance of our microfluidic chip is affected by several factors that include the size and type of tumor cells being captured, body fluid involved (blood vs. non-blood), microfluidic flow rate, design of the chip and its CUs, CTC false positive and capture purity, etc. We have used TU-chip™ for harvesting tumor cells in blood and also, with some modifications to the chip, in non-blood body fluids such as urine [42]. Working with blood and guided by finite element flow modeling and simulations together with experimental results of varying designs of the CU, we found that an arrangement consisting of groups of 3 elliptical micropillars placed in a staggered triangular configuration was not only optimal for CTC capture but also, for an enhanced viability of captured cells. Further, the staggered placement of CUs generates a multi-layered spatial harvesting of CTCs that is clean and debris-free and this greatly enhances the IF identification of CTCs despite the low capture purity of our chip. Further, our TU-chip™ achieved a high CE of >80% which is on par with the 60–92% range of the ISET platform [43–45].

The second issue has 2 parts: the accuracy of the CTC count and the type of cell capture — total vs CTC phenotyping count. Our TU-chip™ uses a label-free, size-based approach to capture the 3 CTC phenotypes in the venous blood of local and metastatic PDAC patients. In contrast, an immunoaffinity-based method uses a specific biomarker or antibody to capture only one CTC phenotype, which typically, is the *E*-CTC. Examples of such an approach include the popular FDA approved CellSearch system [46], NanoVelcro CTC chip [47] and Herringbone HB-chip [48] as they exploit the EpCAM expression for a successful CTC capture. Although they are considered to yield the total CTC count, in reality, they capture a significantly reduced portion of the total CTC population, typically <50% with the *M* and *H* CTCs missing. This explains why the CTC yield of EpCAM-based chips is characteristically low [8,10], which in turn, generates a low CTC detection rate in patients.

Another situation that can lead to an inadvertent reduced CTC count is to use a phenotype-specific biomarker such as cytokeratin to identify and unintentionally select only the *E* CTCs instead of all 3 CTC phenotypes in a size-based platform for CTC capture. Once again, the yield is a subset of the total captured CTC population. Both approaches, using an affinity-based method or an incorrect marker can lead to a greatly reduced CTC capture, which will invariably affect the accuracy of the model. In our work, we employed a size-based approach for a complete cell phenotype capture and used epithelial and mesenchymal markers to properly and fully identify all tumor cells for an accurate total CTC count.

The second part of the issue deals with the use of the total vs phenotype CTC count in the metastatic model. Depending on the intended application, a CTC blood test based on the total CTC count could be useful;

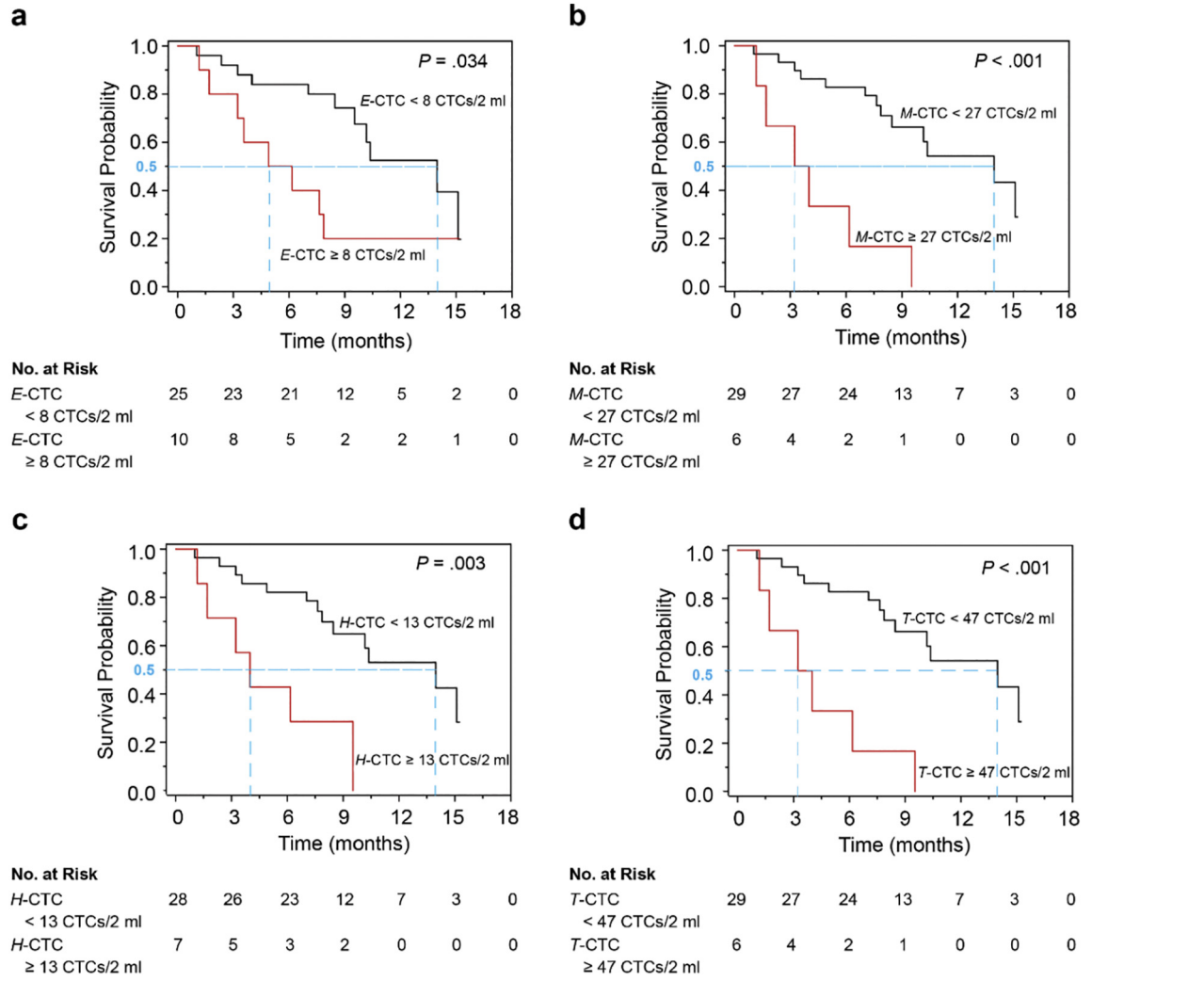

Fig. 7. Unadjusted Kaplan-Meier RFS curves for PDAC patients ($n = 35$) and separated into 2 groups for (a) *E*-CTC, for patients with *E*-CTCs/2 ml < 8 vs ≥8, and the median RFS was 13.97 months (95% CI 8.66–19.28) vs 4.90 months (95% CI 0.87–8.93). (b) *M*-CTC, for patients with *M*-CTCs/2 ml <27 vs ≥27, the median RFS was 13.97 months (95% CI 8.22–19.72) vs 3.23 months (95% CI 0.47–5.99). (c) *H*-CTC, for patients with *H*-CTCs/2 ml < 13 vs ≥13, the median RFS was 13.97 months (95% CI 8.15–19.79) vs 4.00 months (95% CI 2.02–5.98). (d) *T*-CTC, for patients with *T*-CTCs/2 ml < 47 vs ≥47, the median RFS was 13.97 months (95% CI 8.22–19.77) vs 3.23 months (95% CI 0.47–5.99). *P* values are for the log-rank test between survival curves.

however, segregating CTCs into their 3 distinct phenotypes recognizes the unique role played by each CTC phenotype in the cancer pathogenesis, particularly, in tumor metastasis. Changes in the cell phenotype from the highly polarized immotile *E*-CTC in the epithelial state into the motile, cell-death resistive *M*-CTC in the mesenchymal state during an EMT program indicate the beginning of the metastatic process [49,50]. Unlike the immotile *E*-CTC, the *M*-CTC is an effective circulating transporter in the peripheral blood system to allow an extensive infiltration of CTCs throughout the body to form distal tumors. To complete the metastatic process, the reverse change from *M*-CTC back to *E*-CTC via an MET program is a prerequisite for tumor progression and metastasis [51,52]. More importantly, during the forward and reverse transitions between epithelial and mesenchymal states, *H* CTCs are generated and due to their inherent instability, they exhibit increased stemness [53–56]. However, there is increasing evidence that the hybrid phenotype can be stable or at least, metastable. Mathematical modeling works had revealed that transcription factors such as OVOL [57], GRHL2 and microRNA miR-145 [58] act as phenotypic stability factors that enabled cells to have a stable hybrid phenotype and in terms of experimental works, knockdowns of such phenotypic stability factors induced complete EMT process into the mesenchymal phenotype [58,59]. Another potential way a stable hybrid phenotype can be generated is by balancing opposite EMT signalling pathway via a co-treatment of TGFβ and retinoic acid [60] whereby the former induces EMT [61] and the latter inhibits EMT [62]. Besides being a stable state, the stemness exhibited by this phenotype is increased [53–56]. Thus, they are responsible for the intratumoral heterogeneity and tumor infiltration, and are more tumorigenic than either *E* or *M* CTCs [9,15,63,64]. In the univariable CTC count model in Table 4, the *H*-CTC is flagged as the cutoff metric for regulating the metastatic differentiation and this outcome is consistent with the invasiveness nature of hybrid CTCs.

By differentiating the captured CTCs into their distinct phenotypes, we hypothesized that we are able to more accurately and more reliably predict tumor metastasis by monitoring stem cell-like CTCs relative to the immotile and motile tumor cells, that is, the *H*-CTC versus the *E*-CTC and *M*-CTC. We think that this phenotypic profiling constitutes a superior approach than an indiscriminate use of CTCs or a blind use of the total CTC count to predict tumor metastasis.

The third issue is to consider is whether to adopt a univariable vs multivariable CTC phenotype count model in the statistical analysis. To address this issue, we examined the performance of both techniques (Table 4) and the results clearly showed that the disjunctively combined CTC phenotype of '*H*-CTC ≥ 15.0 CTCs/2 ml OR *E*-CTC ≥ 11.0 CTCs/2 ml' generated a better outcome than the univariable model. Our results showed that the multivariable model was able to correctly differentiate all metastatic patients missed-out by the univariable model. Choosing the combined CTC phenotypes, we investigated two multivariable models and found that the mROC model performed better than the commonly used mBLR model for discriminating local and metastatic patients (Table 4 vs Table S5). Further, our mROC analysis showed that the disjunctive ("OR") combination produced better differentiating outcomes than the conjunctive ("AND") combination (Table 4 vs Table S8). In terms of the reliability, the Cohen's kappa of 0.788 showed that the disjunctive combination yielded an outcome that was in substantial agreement with a standard pathology report [65]. On the other hand, the reproducibility of this result was ensured by utilizing the cross-validation technique in validating our models [66]. Lastly, the test results were reported in units of "CTCs per 2 ml" without any manipulation to the blood volume to avoid invoking the assumption that CTCs are evenly distributed in the blood.

Similar to the metastatic model, we compared the accuracy of using the total versus the phenotype CTC count for OS and RFS predictions. Since there is a strong correlation between CTC numbers and poorer survival [11,46,67], our total count is defined by $T = E + M + H$-CTC, in contrast to the traditional concept of the "total" count defined by $T_2 = E + H$-CTC (EpCAM+ or cytokeratin+). However, there is increasing scientific evidence to suggest that the EMT status of CTCs yields a better predictor of disease progression and survival [9,12,68]. In particular, for pancreatic cancer Poruk et al. [9] showed that *E*-CTCs (cytokeratin+) resulted in a poorer survival. Employing a multivariable analysis, they showed that the *E*-CTCs constituted a significant independent predictor of survival. Interestingly, the *H*-CTC was not a significant independent predictor of the OS and we hypothesized that this is because the *H*-CTC count does not accurately represent the tumor burden. The view of metastatic tumor cannot occur unless the epithelial phenotype in generated from the intermediate *H*-CTC via the MET process [52,69] is in line with this hypothesis.

In our work, we employed 3 indices: concordance (C-index) [70], R^2 value [71] and Akaike information criterion (AIC) [72] to assess the relative quality of our phenotype CTC count against the total CTC count. As depicted in Table 5, the result of the 3 indices (higher C-index, higher R^2 value and lower AIC) showed that the phenotype CTC count produced a better OS prognostic model than the T_2 total CTC count. Further, all types of CTCs were significantly associated with recurrence in our study, but none of them were found to be a significant independent predictor of RFS.

Interestingly, although the cutoff points of the metastatic and OS analyses were independently determined, they turned out to have the same discriminant value: *E* ≥ 11.0 CTCs/2 ml for differentiating between metastatic and local tumors, and *E* ≥ 11.0 CTCs/2 ml for differentiating the prognostic OS median of either 5.53 months for metastatic patients or 16.53 months for local patients. Combining these outcomes, it indicates that the cutoff point, *E* ≥ 11.0 CTCs/2 ml not only was able to discriminate between metastatic and local tumors, but also, served to simultaneously predict an OS prognosis for these 2 groups of patients. Note that the OS prognostic prediction of 16.53 months only holds if the condition of the local patient remains unchanged after the blood-draw, which includes undergoing a surgical resection to ensure the cancer does not spread.

In conclusion, by taking into account the specific characteristics of each CTC phenotype, we argued that a disjunctive combination of *E*/*H*-CTC is a better predictor of metastasis and that *E*-CTC is a better predictor of OS with the two serendipitously sharing the same cutoff point. Our work suggested that a phenotype CTC-based blood test has the potential to be developed into an accurate and reliable preoperative prediction of local and metastatic PDAC tumors to complement the traditional imaging tools for tumor staging and surgical resectability. Further, our prefatory study indicated that such an approach can also be used for the OS prognosis of pancreatic cancer patients.

Acknowledgment

We would like to thank the participants and staff of Peking Union Medical College Hospital, and in particular, to Rongrong Zhu, MD for her advice and assistance in experiment management. We would also like to convey our gratitude to CapitalBio Corporation Micro-System Lab in Beijing for their soft lithography support.

Funding sources

National Natural Science Foundation of China, Grant No. 81570061, Key Project of the Science and Technology Program of Zhejiang, Grant No: 2014C03028 and China Scholarship Council (via a scholarship to K.S.C.).

Authors' contributions

R.P.S.H. and M.D. conceived and designed the study with input from Y.S., K.S.C. and G.W. Further, Y.S., A.C., Z.T. and S.C. designed and fabricated the TU-chipTM and developed the microfluidic system setup. Y.

S., K.H.N., A.C., performed the chip characterization and cell line validation experiments. G.W. and M.D. recruited patients and collected all the information from hospital. Y.S. and G.W. conducted the circulating tumor cell assaying experiments. K.S.C. and Y.S. carried out the data and multivariate statistical analyses, data interpretations and generated all graphics. P.F.L. providing special expertise and collaboration in data analysis. R.P.S.H., K.S.C. and Y.S. drafted the manuscript and figures with help from all coauthors. All authors read and approved the final manuscript.

Declaration of Competing Interest

The authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.07.044.

References

[1] Hezel AF, Kimmelman AC, Stanger BZ, Bardeesy N, Depinho RA. Genetics and biology of pancreatic ductal adenocarcinoma. Genes Dev 2006;20(10):1218–49.
[2] Karmazanovsky G, Fedorov V, Kubyshkin V, Kotchatkov A. Pancreatic head cancer: accuracy of CT in determination of resectability. Abdom Imaging 2005;30(4):488–500.
[3] Vincent A, Herman J, Schulick R, Hruban RH, Goggins M. Pancreatic cancer. Lancet 2011;378(9791):607–20.
[4] Somers I, Bipat S. Contrast-enhanced CT in determining resectability in patients with pancreatic carcinoma: a meta-analysis of the positive predictive values of CT. Eur Radiol 2017;27(8):3408–35.
[5] Callery MP, Chang KJ, Fishman EK, Talamonti MS, Traverso LW, Linehan DC. Pretreatment assessment of resectable and borderline resectable pancreatic cancer: expert consensus statement. Ann Surg Oncol 2009;16(7):1727–33.
[6] Diehl SJ, Lehmann KJ, Sadick M, Lachmann R, Georgi M. Pancreatic cancer: value of dual-phase helical CT in assessing resectability. Radiology 1998;206(2):373–8.
[7] Tien YW, Kuo HC, Ho BI, et al. A high circulating tumor cell count in portal vein predicts liver metastasis from periampullary or pancreatic cancer: a high portal venous CTC count predicts liver metastases. Med (Baltimore) 2016;95(16):e3407.
[8] Ankeny JS, Court CM, Hou S, et al. Circulating tumour cells as a biomarker for diagnosis and staging in pancreatic cancer. Br J Cancer 2016;114(12):1367–75.
[9] Poruk KE, Valero V, Saunders T, et al. Circulating tumor cell phenotype predicts recurrence and survival in pancreatic adenocarcinoma. Ann Surg 2016;264(6):1073–81.
[10] Court CM, Ankeny JS, Sho S, et al. Circulating tumor cells predict occult metastatic disease and prognosis in pancreatic cancer. Ann Surg Oncol 2018;25(4):1000–8.
[11] Effenberger KE, Schroeder C, Hanssen A, et al. Improved risk stratification by circulating tumor cell counts in pancreatic cancer. Clin Cancer Res 2018;24(12):2844–50.
[12] Yu M, Bardia A, Wittner BS, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science 2013;339(6119):580–4.
[13] Yang J, Weinberg RA. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. Dev Cell 2008;14(6):818–29.
[14] Lecharpentier A, Vielh P, Perez-Moreno P, Planchard D, Soria JC, Farace F. Detection of circulating tumour cells with a hybrid (epithelial/mesenchymal) phenotype in patients with metastatic non-small cell lung cancer. Br J Cancer 2011;105(9):1338–41.
[15] Jolly MK, Boareto M, Huang B, et al. Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. Front Oncol 2015;5:155.
[16] Chen S, Sun Y, Neoh KH, et al. Microfluidic assay of circulating endothelial cells in coronary artery disease patients with angina pectoris. PLoS One 2017;12(7):e0181249.
[17] Arkin CF, Wachtel MS. How many patients are necessary to assess test performance? JAMA 1990;263(2):275–8.
[18] Kamande JW, Hupert ML, Witek MA, et al. Modular microsystem for the isolation, enumeration, and phenotyping of circulating tumor cells in patients with pancreatic cancer. Anal Chem 2013;85(19):9092–100.
[19] Grimes DA, Schulz KF. Compared to what? Finding controls for case-control studies. Lancet 2005;365(9468):1429–33.
[20] Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005;4(4):287–91.
[21] Dicato M, Plawny L, Diederich M. Anemia in cancer. Ann Oncol 2010;21(Suppl. 7)(vii167–72).
[22] Amin MB, Edge S, Greene F, et al. AJCC Cancer Staging Manual. . 8 ed.Springer International Publishing; 2017.
[23] Nijkamp MM, Span PN, Hoogsteen IJ, van der Kogel AJ, Kaanders JH, Bussink J. Expression of E-cadherin and vimentin correlates with metastasis formation in head and neck squamous cell carcinoma patients. Radiother Oncol 2011;99(3):344–8.
[24] Yamada S, Fuchs BC, Fujii T, et al. Epithelial-to-mesenchymal transition predicts prognosis of pancreatic cancer. Surgery 2013;154(5):946–54.
[25] Scheibelhofer O, Besenhard MO, Piller M, Khinast JG. Comparing particle size distributions of an arbitrary shape. Powder Technol 2016;294:134–45.
[26] Shultz EK. Multivariate receiver-operating characteristic curve analysis - prostate-cancer screening as an example. Clin Chem 1995;41(8):1248–55.
[27] Yadav S, Shukla S. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. 2016 IEEE 6th International conference on advanced computing (IACC); 2016. p. 78–83.
[28] Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res 2004;10(21):7252–9.
[29] Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and cox regression. Am J Epidemiol 2007;165(6):710–8.
[30] Vona G, Sabile A, Louha M, et al. Isolation by size of epithelial tumor cells : a new method for the immunomorphological and molecular characterization of circulatingtumor cells. Am J Pathol 2000;156(1):57–63.
[31] Mohamed H, Murray M, Turner JN, Caggana M. Isolation of tumor cells using size and deformation. J Chromatogr A 2009;1216(47):8289–95.
[32] Strony J, Beaudoin A, Brands D, Adelman B. Analysis of shear stress and hemodynamic factors in a model of coronary artery stenosis and thrombosis. Am J Physiol 1993;265(5 Pt 2):H1787–96.
[33] Zhao S, Chen C, Chang K, et al. CD44 expression level and isoform contributes to pancreatic cancer cell plasticity, invasiveness, and response to therapy. Clin Cancer Res 2016;22(22):5592–604.
[34] Rice AJ, Cortes E, Lachowski D, et al. Matrix stiffness induces epithelial-mesenchymal transition and promotes chemoresistance in pancreatic cancer cells. Oncogenesis 2017;6(7):e352.
[35] Kim Y, Han D, Min H, Jin J, Yi EC, Kim Y. Comparative proteomic profiling of pancreatic ductal adenocarcinoma cell lines. Mol Cell 2014;37(12):888–98.
[36] Aiello NM, Maddipati R, Norgard RJ, et al. EMT subtype influences epithelial plasticity and mode of cell migration. Dev Cell 2018;45(6):681–95 [e4].
[37] Nordgard O, Tjensvoll K, Gilje B, Soreide K. Circulating tumour cells and DNA as liquid biopsies in gastrointestinal cancer. Br J Surg 2018;105(2):e20–110.
[38] Ferreira MM, Ramani VC, Jeffrey SS. Circulating tumor cell technologies. Mol Oncol 2016;10(3):374–94.
[39] Satelli A, Mitra A, Cutrera JJ, et al. Universal marker and detection tool for human sarcoma circulating tumor cells. Cancer Res 2014;74(6):1645–50.
[40] Castle J, Morris K, Pritchard S, Kirwan CC. Challenges in enumeration of CTCs in breast cancer using techniques independent of cytokeratin expression. PLoS One 2017;12(4):e0175647.
[41] Yoo W, Mayberry R, Bae S, Singh K, Peter He Q, Lillard Jr JW. A study of effects of multicollinearity in the multivariable analysis. Int J Appl Sci Technol 2014;4(5):9–19.
[42] Chen A, Fu G, Xu Z, et al. Detection of urothelial bladder carcinoma via microfluidic immunoassay and single-cell DNA copy-number alteration analysis of captured urinary-exfoliated tumor cells. Cancer Res 2018;78(14):4073–85.
[43] Lin HK, Zheng S, Williams AJ, et al. Portable filter-based microdevice for detection and characterization of circulating tumor cells. Clin Cancer Res 2010;16(20):5011–8.
[44] Chen F, Wang S, Fang Y, et al. Feasibility of a novel one-stop ISET device to capture CTCs and its clinical application. Oncotarget 2017;8(2):3029–41.
[45] Sun N, Li X, Wang Z, Li Y, Pei R. High-purity capture of CTCs based on micro-beads enhanced isolation by size of epithelial tumor cells (ISET) method. Biosens Bioelectron 2018;102:157–63.
[46] Cristofanilli M, Budd GT, Ellis MJ, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med 2004;351(8):781–91.
[47] Lu YT, Zhao L, Shen Q, et al. NanoVelcro Chip for CTC enumeration in prostate cancer patients. Methods 2013;64(2):144–52.
[48] Stott SL, Hsu CH, Tsukrov DI, et al. Isolation of circulating tumor cells using a microvortex-generating herringbone-chip. Proc Natl Acad Sci U S A 2010;107(43):18392–7.
[49] Krebs AM, Mitschke J, Lasierra Losada M, et al. The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. Nat Cell Biol 2017;19(5):518–29.
[50] Rhim AD, Mirek ET, Aiello NM, et al. EMT and dissemination precede pancreatic tumor formation. Cell 2012;148(1–2):349–61.
[51] Jayachandran A, Dhungel B, Steel JC. Epithelial-to-mesenchymal plasticity of cancer stem cells: therapeutic targets in hepatocellular carcinoma. J Hematol Oncol 2016;9(1):74.
[52] Ikeda S, Schwaederle M, Mohindra M, Fontes Jardim DL, Kurzrock R. MET alterations detected in blood-derived circulating tumor DNA correlate with bone metastases and poor prognosis. J Hematol Oncol 2018;11(1):76.
[53] Bierie B, Pierce SE, Kroeger C, et al. Integrin-beta4 identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells. Proc Natl Acad Sci U S A 2017;114(12):E46-2337.
[54] Andriani F, Bertolini G, Facchinetti F, et al. Conversion to stem-cell state in response to microenvironmental cues is regulated by balance between epithelial and mesenchymal features in lung cancer cells. Mol Oncol 2016;10(2):253–71.
[55] Jolly MK, Jia D, Boareto M, et al. Coupling the modules of EMT and stemness: a tunable 'stemness window' model. Oncotarget 2015;6(28):25161–74.
[56] Strauss R, Li ZY, Liu Y, et al. Analysis of epithelial and mesenchymal markers in ovarian cancer reveals phenotypic heterogeneity and plasticity. PLoS One 2011;6(1):e16186.
[57] Jia D, Jolly MK, Boareto M, et al. OVOL guides the epithelial-hybrid-mesenchymal transition. Oncotarget 2015;6(17):15436–48.
[58] Jolly MK, Tripathi SC, Jia D, et al. Stability of the hybrid epithelial/mesenchymal phenotype. Oncotarget 2016;7(19):27067–84.

[59] Watanabe K, Villarreal-Ponce A, Sun P, et al. Mammary morphogenesis and regeneration require the inhibition of EMT at terminal end buds by Ovol2 transcriptional repressor. Dev Cell 2014;29(1):59–74.
[60] Biddle A, Gammon L, Liang X, Costea DE, Mackenzie IC. Phenotypic plasticity determines cancer stem cell therapeutic resistance in oral squamous cell carcinoma. EBioMedicine 2016;4:138–45.
[61] Biddle A, Liang X, Gammon L, et al. Cancer stem cells in squamous cell carcinoma switch between two distinct phenotypes that are preferentially migratory or proliferative. Cancer Res 2011;71(15):5317–26.
[62] Metallo CM, Ji L, de Pablo JJ, Palecek SP. Retinoic acid and bone morphogenetic protein signaling synergize to efficiently direct epithelial differentiation of human embryonic stem cells. Stem Cells 2008;26(2):372–80.
[63] Zhang S, Wu T, Peng X, et al. Mesenchymal phenotype of circulating tumor cells is associated with distant metastasis in breast cancer patients. Cancer Manag Res 2017;9:691–700.
[64] Jolly MK, Ware KE, Gilja S, Somarelli JA, Levine H. EMT and MET: necessary or permissive for metastasis? Mol Oncol 2017;11(7):755–69.
[65] Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005;37(5):360–3.
[66] Koul A, Becchio C, Cavallo A. Cross-validation approaches for replicability in psychology. Front Psychol 2018;9:1117.
[67] Hofman V, Bonnetaud C, Ilie MI, et al. Preoperative circulating tumor cell detection using the isolation by size of epithelial tumor cell method for patients with lung cancer is a new prognostic biomarker. Clin Cancer Res 2011;17(4):827–35.
[68] Satelli A, Mitra A, Brownlee Z, et al. Epithelial-mesenchymal transitioned circulating tumor cells capture for detecting tumor progression. Clin Cancer Res 2015;21(4):899–906.
[69] Nieto MA, Huang RY, Jackson RA, Thiery JP. Emt: 2016. Cell 2016;166(1):21–45.
[70] Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 2011;30(10):1105–17.
[71] Korn EL, Simon R. Measures of explained variation for survival data. Stat Med 1990;9(5):487–503.
[72] Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control 1974;19(6):716–23.