Enumeration, characterisation and clinicopathological significance of circulating tumour cells in patients with colorectal carcinoma

Faysal Bin Hamid[1], MSc.

Cu-Tai Lu[2], MBBS, MMed, FRACS.

Marco Matos[3], RACP.

Tracie Cheng[1], BSc (Hon)

Vinod Gopalan[1*], PhD

Alfred King-yin Lam[1,4*] MBBS, PhD, FRCPA.

[1]Cancer Molecular Pathology, School of Medicine, Griffith University, Gold Coast, Queensland 4222, Australia.

[2]Colorectal Surgery, Gold Coast University Hospital, Gold Coast, Australia.

[3]Oncology, Gold Coast University Hospital, Gold Coast, Australia.

[4]Pathology Queensland, Gold Coast University Hospital, Gold Coast, Australia

***Corresponding authors:**

Dr Vinod Gopalan

School of Medicine, Griffith University, Gold Coast QLD 4222, Australia. E-mail: v.gopalan@griffith.edu.au Telephone +61 7 56780717.

Professor Alfred K Lam (for correspondence and reprints request)

Head of Pathology, Griffith Medical School, Gold Coast QLD 4222, Australia. E-mail: a.lam@griffith.edu.au Telephone +61 7 56780718.

Abstract

Background: The purposes of the study were to enumerate and characterise the circulating tumour cell (CTC) and cluster/micro-emboli (CTM) in blood from patients with colorectal carcinoma (CRC) as well as to investigate their clinical relevance.

Methods: Peripheral blood of six healthy donors (control) and sixty-two patients with colorectal carcinoma were collected to isolate CTCs by an immunomagnetic negative selection approach. EPCAM and cytokeratin 18 (CK18) antibodies were used to identify the CTCs. The size and the phenotypic variations were evaluated to characterise these isolated CTCs. Additionally, mRNA expressions of the CTCs and the corresponding primary carcinoma were assessed using a multi-gene panel to determine the cellular heterogeneities between CTCs and primary carcinoma.

Results: We detected CTCs and CTMs in 72% (41/57) and 32% (18/57) of the patients with CRC, respectively. The total number and length were significantly higher ($p<0.0001$) in the CTCs than the CTMs. CTCs, especially $EPCAM^{Positive}CK18^{Pososit ve}$ subclones, were detected more in the patients with advanced pathological cancer stages when compared to those with early cancer stages (mean: 12.5 vs 4.0, $p=0.0068$). mRNA profiling of CTCs unveiled three different CTC subtypes expressing epithelial, epithelium-mesenchymal transition (EMT) and stemness signatures, which were different from those of the primary carcinoma. The expressions of EPCAM, HRAS, BRAF, TP53, SLUG, TWIST1, CD44 and MMP9 of CTCs were altered when compared to the primary tumours in patients with CRC.

Conclusion: Our findings provide insights into the biology of the CTC, presence of heterogeneous CTC populations in CRC and differential expression of genes in different pathological stages of CTC which could improve the management of patients with CRC.

Keywords: Circulating tumour cells, negative selection method, gene expression, colorectal carcinoma.

Introduction

The presence of the circulating tumour cells (CTCs) has been demonstrated as a predictive and prognostic marker in patients with colorectal carcinoma (CRC) [1-3]. Identification of CTCs is generally based on parameters of the cancer cells such as surface antigen, size, elasticity, cellular function etc [3]. The immunomagnetic bead-based methods targeting cellular markers (surface antigens) is a common means to isolate the heterogeneous CTC populations by either positive selection system or negative selection system [1, 2, 4]. The positive selection system (e.g. CellSearch) can isolate specific type of CTCs such as epithelial CTCs with the use of epithelial cell adhesion molecule (EPCAM)-labelled immunomagnetic beads. In contrast, the negative selection system (e.g. RosetteSep^{TM} assays) separates heterogeneous CTCs by removing cells such as leukocytes and platelets using CD45 or CD36-labelled immunomagnetic beads [5, 6].

CTCs in CRCs were often identified by expressions of epithelial and/or mesenchymal markers (EPCAM, cytokeratins 8, 18, 19, 20 or pan-CKs, plastin 3, vimentin) [1, 3]. Immunofluorescence-based studies showed that these markers were expressed heterogeneously in CTCs [7, 8]. Moreover, morphological features such as cell length (>8-10μm) and high nuclear-cytoplasmic ratio have been considered for detection of CTC in prostate adenocarcinoma [9-12]. Nevertheless, the use of these parameters to detect CTC in CRC have not been demonstrated.

Beyond the enumeration and analysing the morphological varieties of CTC, molecular characterisation of the CTCs could shed light on some concealed mechanisms driving cancer invasion and metastasis [13, 14]. During the transformation of the epithelial phenotypes to the mesenchymal phenotypes, CTCs undergo complex and dynamic series of structural and functional changes to achieve increase cell motility and stemness behaviour [15, 16]. In particular, the epithelial markers (EPCAM, CK) are downregulated, and different

transcription factors such as SNAI1, SLUG, TWIST1 are upregulated in CTC, leading to the epithelium-mesenchymal transition (EMT) [15-18]. EMT enables CTCs to abolish the cell-cell interaction and remodel the cytoskeletal network to gain the migratory capacity. Another hallmark feature of EMT, proteolytic degradation of the extracellular matrix, was reported to be induced by the matrix metalloproteinase (MMPs) activity [19].

A few groups have compared the molecular features of CTCs and the corresponding primary tumour which revealed some special modifications in CTCs, but the data are not insufficient to utilize these findings in the clinical settings [20, 21]. In this present study, we aimed to isolate and enumerate heterogeneous populations of CTCs in colorectal carcinoma and study the morphological and phenotypic characteristics of the CTCs. To evaluate the molecular alterations, we have analysed the mRNA expressions of 14 genes in both CTC populations and primary tumour tissue specimens from colorectal carcinoma. In addition, we investigated the correlations among the morphological, phenotypical and molecular features of the heterogeneous populations of CTCs with the pathological stages of the colorectal carcinoma.

Patients and Methods

Cell lines

For spiking experiments, three colon cancer cell lines - SW48, SW-480 and HCT-116 were obtained from American Type Culture Collection (ATCC), and cultured in Roswell Park Memorial Institute culture medium (RPMI 1640) (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% foetal bovine serum (FBS) (Thermo Fisher Scientific), and 1% penicillin-streptomycin (Thermo Fisher Scientific). Upon trypsinisation, the cells were used for spiking and immunofluorescence experiments.

Patients Cohort

Sixty-two patients (38 men, 19 women) with CRC and six healthy individuals were prospectively recruited at Gold Coast University Hospital from May 2017 to June 2019. The mean age of the patients was 67 years. These patients were under the clinical management of the authors (CTL, MM and AKL) from the same clinical team. The demographic information and pathological parameters were obtained from the clinical teams. A preliminary test was done on five patients using six antibodies to identify CTCs in CRC in order to choose the best ones for the study. Ethical approval for this study was obtained from the Griffith University Human Research Ethics Committee (GU Ref No: MSC/17/10/HREC). From each of these patients, 5 ml of peripheral blood was collected in heparin-containing BD (Becton Dickinson, Franklin Lakes, NJ, USA) vacutainer tubes at the time of surgery of CRC.

CTC Enrichment and Isolation

Isolation of CTCs was performed from freshly obtained peripheral blood of the patients with CRC using EasySep™ Direct Human CTC Enrichment Kit (STEMCELL™ Technologies, Vancouver, BC, Canada). Haematopoietic cells and platelets were removed by a cocktail of antibody-labelled immunomagnetic beads targeting CD2, CD14, CD16, CD19,

CD45, CD61, CD66b, and glycophorin A. In brief, 5ml of the whole blood was incubated with "Enrichment cocktail" at room temperature for 5 minutes. Approximately 250µl RapidSphere from the kit was mixed into the sample. The fraction of blood with CTC was collected followed by magnetic removal of the unwanted blood cells. Finally, the enriched CTCs were centrifuged at 450x g relative centrifugal force (rcf) for 7 minutes and resuspended in 200µl CTC medium. The CTC medium comprised of L-glutamine positive Roswell Park Memorial Institute (RPMI) 1640 medium containing 10% foetal bovine serum, 1% penicillin-streptomycin, recombinant human epidermal growth factor (EGF) (20ng/ml) (PeproTech, Rocky Hill, NJ, USA), recombinant human basic FGF (20ng/ml) (PeproTech), recombinant human transforming growth factor (TGF)-β1(20ng/ml) (PeproTech), insulin (20ng/ml) (Thermo Fisher Scientific) and B27 (20ng/ml) (Thermo Fisher Scientific). The enriched CTCs were transferred to 96-well plate (50µl per well) for downstream experiments.

Cell Spiking Experiments

To validate the isolation technique, colon cancer cells from SW-48, SW-480 and HCT-116 cell lines were used. A cell tracker dye, Green CMFDA (5-chloromethylfluorescein diacetate) (Abcam, Cambridge, UK), was used to stain the colon cancer cells. Briefly, the cells were incubated with 10µM CMFDA in serum-free RPMI medium for one hour at 37 °C. After incubation, the cancer cells were washed with phosphate-buffered saline (PBS) twice and disassociated into isolated cells by trypsinisation. The stained cancer cells were counted using fluorescence microscope (Olympus, Tokyo, Japan). Then, 0, ~1, ~10 and ~100 cells were spiked in 5ml of whole blood of the healthy individuals or serum-free RPMI medium. The cells were then recovered using the EasySep™ kit as described in the previous section. Recovered cancer cells were counted under fluorescent microscope (Olympus).

Immunofluorescence Staining

For the immunofluorescence staining, the enriched CTCs were fixed with ice-cold 100% methanol in -20^{0}C for 10 minutes. The cells were washed with 100 µl of PBS and were permeabilized with 0.1% Triton X-100 for 10 minutes. They were then washed with 100 µl of PBS and incubated with bovine serum albumin (BSA), mouse anti-EPCAM (AUA1) antibody (dilution 1:100 in PBS, Thermo Fisher Scientific, Waltham, MA, USA) and goat anti-CK18 antibody (ab219271) (dilution 1:100 in PBS, Abcam, Cambridge, UK) for 2 hours at room temperature. For the training set, we included 4 additional antibodies: goat anti-SNAI1 (E-18) (dilution of 1:100, sc-10432, Santa Cruz Biotechnology, Santa Cruz, CA, USA), mouse anti-Bcl-2 (C-2) (dilution of 1:100, sc-7382, Santa Cruz Biotechnology), mouse anti-E-cadherin (G-10) (dilution of 1:100, sc-8426, Santa Cruz Biotechnology) and goat anti-MMP9 (C-20) (dilution of 1:100, sc-6840, Santa Cruz Biotechnology). After another washing step with PBS, the cells were incubated with goat-anti-mouse IgG fluorescein isothiocyanate (FITC) antibody (dilution of 1:100, Sigma Aldrich, St. Louis, MO, USA) and mouse anti-goat IgG (H+L) Texas Red antibody (dilatation of 1:100, Sigma Aldrich) antibody for one hour. The cells were washed with PBS and incubated with Hoechst 33342 (Thermo Fisher Scientific) to stain the nucleus for 5 minutes. Finally, the cells were washed with PBS and visualized by fluorescent microscope equipped with CellSens Standard software (Olympus). The images of the cells were acquired using 3 filters; Hoechst 33342 filter (excitation: 352, emission: 462), FITC filter (excitation: 490, emission: 525) and Texas red filter (excitation: 596, emission: 615) at 40x magnification. The total nucleated events were counted using ICY, an open community platform for bioimage informatics, providing software resources to visualize, annotate and quantify bioimaging data (http://icy.bioimageanalysis.org). Also, the sizes and fluorescent intensity of the CTCs were measured by the CellSens Standard and Fiji/ImageJ software, respectively.

RNA Extraction and Quantitative Real-Time polymerase chain reaction (qRT PCR)

Total RNA was extracted from the enriched CTC fractions and the tumour specimens using the RNeasy mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instruction. The RNA was eluted in 50µl. Five microliters of RNA were used for complementary DNA (cDNA) synthesis using QuantiTect Reverse Transcription Kit (Qiagen). To remove contaminating genomic DNA, total RNA was incubated by gDNA Wipeout buffer at 42^0C for two minutes. A total volume of 20 µl of cDNA was synthesized at 42^0C at 15 minutes according to the manufacturer's protocol. The resulting cDNAs were diluted in nuclease-free water to a final concentration of 30ng/µl. Primer3 software (http://bioinfo.ut.ee/primer3-0.4.0/primer3/) were used to design and analyse the parameters of the primers**.** The sequences of the primers were listed in supplementary Table 1. Real-time PCR reactions of 14 genes were performed using QuantiTect Syber Green PCR kit (Qiagen) in QuantStudio 6 Flex Real-Time PCR system (Applied Biosystems, Waltham, MA, USA). Briefly, the qPCR reaction was carried out in a total volume of 10ul reaction mixture containing 5µl of 2x QuantiTect Syber Green PCR master mix, 0.5ul of 10uM each primer and 2µl of cDNA (30ng/ul) and 2ul of nuclease-free double-distilled water. Cycling conditions were as following: an initial PCR activation step of 95^0C for 15 minutes, 40 cycles of 94^0C for 15 seconds, 54-62^0C for 30 seconds and 72^0C for 30 seconds. The housekeeping gene, *GAPDH (Glyceraldehyde 3-phosphate dehydrogenase),* was used as an internal control to normalize the PCR reactions. The relative gene expression level was calculated using the delta-delta Ct method, and log10 value of the fold change was plotted in the heatmap.

Statistical Analyses

The statistical analyses were performed using GraphPad Prism Software 5.03 (GraphPad Software Inc., San Diego, CA, USA). Statistical significance was analysed using Mann-Whitney non-parametric t-test. A *p*-value of <0.05 was considered statistically significant. Heatmap was generated with BioVinci software (Bioturing Inc., San Diego, CA, USA) using the Euclidean distance and Ward's minimum clustering method. The values were estimated from the log10 value of the relative quantification (RQ) of each gene.

Results

Validation of Negative Selection via Spiking Method

Using the negative selection (NS) method, the spiked-in cells were recovered from the blood of healthy persons (Fig. 1A). Approximately 65% and 70% of SW-48 cells and SW-480 cells were recovered respectively from 100 spiked cells (Fig. 1B). The HCT-116 cells showed the highest recovery rate of 72%. In addition, when ten cells were spiked-in, about 50% of cells were captured from all the cell lines. No cell was recovered when a single cell was spiked (Fig. 1B).

Enumeration of CTCs in patients with CRC

Initially, a panel of 6 antibodies was used for screening of CTCs from 5 patients with CRC (Fig. 2A). The EPCAM and CK18 proteins were expressed in CTCs from all these selected cases. They were selected for further validation of CTC validation for the 57 patients with CRC. The expression levels of EPCAM and CK18 were also investigated in the three colon cancer cell lines (Fig. 2B). They were slightly lower in SW-48 cells when compared to SW-480 cells or HCT-116 cells.

CTCs were detected in 72% (41 of 57) of patients with CRC (Fig. 2C). In addition, we identified CTC clusters or micro emboli (CTMs) from 32% (18/57) of the patients with CRC (Fig. 2C). A wide range of number (range: 0-88) of CTCs were noted in patients with CRC whereas no CTC was detected in the healthy donors (HD) (Fig. 2D). The number of CTCs were observed significantly higher than the CTMs ($p=0.0001$) (Fig. 2E). Besides, many white blood cells were detected in both healthy donors (mean: 10,096.0) and patients with CRC (mean: 11,267.7) (Suppl. Fig. 1).

Morphological and Phenotypical Characterization of CTCs

We have observed morphological heterogeneities of the CTCs in different patients with CRC (Fig. 3A). The diameter of the cells in CTMs was significantly smaller than the individual CTCs (Fig. 3B). The mean length of the CTCs was 29.57 μm (range: 17.63-44.77 μm) whereas that of CTMs was 14.86 μm (range 9.95-24.10 μm).

In this study, CTCs were categorised based on expressions of EPCAM and CK18 proteins into 3 classes: $EPCAM^{Pos}CK18^{Pos}$, $EPCAM^{Pos}CK18^{Neg}$ and $EPCAM^{Neg}CK18^{Pos}$ subclones (Fig. 3C). $EPCAM^{Pos}CK18^{Pos}$ subclones were found in 92.7% (38 of 41) of CTC positive cases (mean: 7.4, range: 0-60). $EPCAM^{Pos}CK18^{Neg}$ and $EPCAM^{Neg}CK18^{Pos}$ subpopulations were detected in 65.9% (mean: 2.0, range: 0-18) and 24.6% (mean: 1.29, range: 0-12) of patients with CRC respectively (Fig. 3D).

Gene Profiling of The CTC Fraction and Tumour

To evaluate the molecular differences between the CTCs and the matched primary tumour samples, we performed qPCR to study a gene panel of 14 genes. The gene-set was comprised of epithelial genes (*EPCAM*, *CK20*), oncogenes (*KRAS*, *NRAS*, *HRAS*, *BRAF*), tumour suppressor genes (*TP53*, *FOXO3*), Epithelial-mesenchymal transition (EMT) genes (*SNAI1*, *SLUG*, *TWIST1*), stemness genes (*CD133*, *CD44*) and extracellular matrix (ECM) degrading gene (*MMP9*). We analysed the data with unsupervised hierarchical clustering (average clustering method). The gene clustering revealed different expression patterns of 14 genes between the CTC fractions and the primary tumour (Fig. 4).

The CTCs were apparently grouped into three patient clusters: cluster 1, cluster 2 and cluster 3 (Fig. 4A). In cluster 1, epithelial markers (*EPCAM, CK20*) were highly expressed in 90.9%, and at least two oncogenes and one tumour suppressor gene were expressed in 50% and 63.6% of the patients respectively. Moreover, at least two of the EMT and stem cell

markers were positively detected in 36.4% and 40.9%, while *MMP9* expression was low in 31.8% of the patients. In cluster 2, slightly lower expressions (57.9%) of the epithelial genes were found compared to the cluster 1. A similar pattern was also observed in oncogenes (42.10%) and tumour suppressor genes (57.9%). The expression levels of two EMT genes, *SNAI1* and *SLUG*, as well as stemness markers were higher in 52.6%, but lower level of *MMP9* was expressed in 79% of the cases. In cluster 3, most of the patients had lower expression of epithelial and tumour suppressor markers (62.5% and 37.5% respectively), and the oncogenes were overexpressed in 75% of the patients. Also, a complex relationship was observed among the EMT, stemness markers and *MMP9* in cluster 3, i.e. the latter two types of markers were higher when EMT markers were lower and vice versa.

Gene expression of the tumour tissues could be assembled into 2 clusters (Fig. 4B). In cluster 1, most of the tumour tissues overexpressed epithelial (82.8%) and tumour suppressor genes (58.6%). Although oncogenes were often detected, the expression of *KRAS* and *HRAS* was antagonistic. EMT and stem cell markers were slightly lower in most of the cases. In cluster 2, we observed lower expression of the epithelial, oncogenes and tumour suppressor markers comparing to cluster 1, but stemness and ECM degrading marker *MMP9* was overexpressed (71.4% and 85.7% respectively).

Compared to the primary tumour samples, CTCs expressed higher level ($p<0.0001$) of epithelial marker (*EPCAM*) (Fig. 5). Among the oncogenes, *HRAS* showed significant overexpression ($p< 0.0001$) whereas *BRAF* expression was noted to be reduced ($p=0.0096$) in the CTC population. Although overexpression of EMT genes (*SLUG* and *TWIST1*) were observed, tumour suppressor genes were suppressed. Among the stemness related genes, *CD44* expression was significantly lower ($p< 0.05$) in the CTCs when compared to the primary carcinomas. On the other hand, *MMP9*, an ECM degrading gene, was highly expressed in tumour tissues compared to CTCs ($p<0.0001$).

Clinical and Pathological Correlations

The presence of CTC were significantly correlated with tumour grade, lymph nodal status and pathological stages in patients with CRC (Table 2). However, the presence of CRC they did not show any association with the age, gender of the patients with CRC or size, grade of the CRC.

We classified the CRC into early pathological stages (stage I or II, n=37) and advanced pathological stages (III orIV, n=20) as in the previous report [22]. The number of CTCs was higher (mean: 20.0 vs 6.135, p=0.0019) in patients with CRC of advanced pathological stages than those with CRC of early pathological stages (Fig. 6A). Two or more CTCs were detected in 21 of 37 patients with early pathological stages but all except one of the 20 patients with advanced pathological stages. In addition, 11 patients with advanced pathological staged cancer had ≥10 CTCs (range: 10-88). Similarly, the number of CTMs were observed more (mean: 2.368 vs 1.60, p=0.721) in advanced pathological stages (Fig. 6B). However, the difference did not reach statistical significance. Also, $EPCAM^{pos}CK18^{Pos}$ CTCs were significantly higher (mean: 12.5 vs 4.0, p=0.0068) in patients with advanced pathological stages of CRC (Fig. 6C). There was no significant difference (mean: 3.04 vs 1.33, p=0.1125) in $EPCAM^{Pos}CK18^{Neg}$ subtype of CTCs between these two groups (Fig. 6D). Similarly, no assocation (mean: 1.05 vs 1.79, p=0.068) was observed between $EPCAM^{Neg}CK18^{Pos}$ subtype of CTCs and pathological stages (Fig. 6E). In additon, larger CTCs were observed in patients having CRC with advanced pathological stages when compared to those with early pathological stages (mean: 27.60 vs 26.04, p=0.036) (Fig. 6F). Furthermore, the expression of stemness markers in CTCs were higher in the CRC of early pathological stages than the CRC of advanced stages (Fig. 6G).

Discussion

Over the past two decades, numerous studies have reported isolating CTCs from patients with cancers, including colorectal cancers [2, 3, 7, 8]. Lack of a reliable approach for identification of CTC is a barrier for its use as a clinical biomarker. One of the greatest challenges in isolation of CTC is the small number of CTC and short lifespan of CTC in the blood which hinders the capture of viable CTC as well as for use in molecular studies. Moreover, CTCs can escape detection because of its heterogeneity in protein expression. For instance, the CellSearch system implements positive immunomagnetic separation, which can capture only $EpCAM^{+}$ $CK^{+}CD45^{-}$ cells with an efficiency rate of 36.8-43.2% [7, 23]. Studies showed that CTCs are often more divergent than it was thought [1, 3, 9]. Negative selection assays used to isolate CTC were designed to remove the high concentration of the peripheral blood cells. Ozkumur *et al*. compared the positive selection assay and negative selection assay using spike-in experiments found a higher efficacy (>85%) in the isolation of the mesenchymal CTCs by the negative selection technique [24]. In this study, we showed that negative selection approach could isolate heterogeneous CTC populations from >70% of the patients with CRC. Additionally, the identification of $EPCAM^{Pos}CK18^{Pos}$, $EPCAM^{Pos}CK18^{Neg}$ and $EPCAM^{Neg}CK18^{Pos}$ subclones indicated the presence of heterogeneous CTC phenotypes in the patients with CRC.

Studies reported that CTC could exhibit both epithelial and mesenchymal phenotypes [8, 25-27]. The present study focused on the detection of CTCs and CTMs based on epithelial markers. It demonstrated variable expression levels of EPCAM and CK18 and confirmed the presence of different CTC phenotypes in CRC. For instance, $EPCAM^{Pos}CK18^{Pos}$ subclones indicated epithelial phenotype whereas $EPCAM^{Pos}CK18^{Neg}$ and $EPCAM^{Neg}CK18^{Pos}$ subclones indicated EMT phenotypes of CTCs. Zhao *et al*. demonstrated that the number of EMT type CTCs were higher than epithelial type CTCs and

associated with tumour stages, lymph nodal status and distant metastasis [28]. On the other hand, Wu and his colleagues reported that epithelial type CTCs were more common than EMT type CTCs, which is consistent with our result [27]. In addition, Zhang *et al*. reported that higher CTC count and presence of CTM was associated with worse prognosis of patients with CRC [29]. Furthermore, we noted a significant correlation between CTC counts and pathological stages in patients with CRCs. However, no clinical correlations were observed in EMT subclones of CTCs or CTMs. Further studies are warranted to investigate the reason.

Morphological features can characterise different types of CTCs. The size of the CTCs was larger (>8μm) than the blood cells such as lymphocytes and neutrophils supporting its potential use in isolation of CTCs [30]. As a proof-of-concept, different filtration methods could isolate the CTCs based on the cellular size [30]. Our results demonstrated that CTCs could be different in size (in term of their diameters) and smaller than cancer cells in established cell lines [31]. Likewise, the size of the cells present in CTM was found to be smaller than those of individual CTC [32]. It is likely that tumour cells within CTM maintain smaller size for them to traverse through the narrow blood capillaries [33]. Also, we have found an association between the size of the CTCs and the pathological stages in CRC. Meng and colleagues have reported that explicit correlations between CTC's diameter and patients with latent, metastatic and non-metastatic cancers [34]. As cell length of CTC is difficult to determine the biological significance in metastasis, more extensive studies are needed to validate the results.

The cell number in a CTM or size of the CTMs may have biological and clinical significances in patients with cancer. For example, larger CTM was associated with worse prognosis of patients with metastatic breast carcinoma [32]. In another study, neither CTM count nor CTM size correlated with the pathological stages in the patients with CRC [35]. Our results could not demonstrate any relationship of the cell number of CTMs with

pathological staging in patients with CRC. Decrease of the cell number of the CTMs may be resulted from the dissociation of the individual cell from the clusters due to cellular collisions (e.g. erythrocyte, leucocyte, macrophage etc.), shear stresses or merely handling [36]. Further investigation is warranted to fully understand the biological characteristics of CTMs in patients with CRCs.

Besides enumeration and biological features of CTC, molecular characterisation of the CTCs can be a platform to study the cellular heterogeneities, resistance mechanisms and therapeutic targets in cancer. The heterogeneity among the CTCs could explain the development of resistance to systemic chemotherapy and consequently sub-clonal evolution. Our data indicated that primary tumour tissue had more diverse and different expression patterns when compared with CTCs. This is likely due to the presence of significant numbers of non-cancer cells in primary tumour tissue. Tumour cells undergo EMT to be released from the primary tumour [15]. EMT, the transient epithelial to mesenchymal stage, can be induced by the *RAS* oncogene and the stemness markers such as *CD133* and *CD44* [37-39]. *NRAS* and *KRAS*/*HRAS* are required to the cell adhesion and motility, respectively, whereas *TP53* counteracts of RAS function [40]. The alteration of these genes impedes their regular "molecular switch" functions and turns on aberrant growth, proliferation, and migration and consequently oncogenesis. In addition, CD44 regulates the cytoskeletal remodelling by the transcription of *MMPs* and nuclear translocation of the SNAI1/β-catenin complex [41]. Furthermore, Dhar *et al*. demonstrated that MMP activity of the CTCs was increased 200-fold, which facilitates them to dislodge from the primary tumour [42]. Our findings suggested that heterogeneous CTCs were classified into 3 clusters. Different gene expression profiling of the CTCs revealed the inter-patient heterogeneity, which highlighted distinctive mechanisms *en route* to the metastasis among the patients with CRC [43, 44]. We found a subtype of CTCs expressing higher epithelial genes *EPCAM* and *CK20* representing the

epithelial CTC subclones (cluster 1). It is worth noting that *EPCAM* regulates the Ras/Raf/ERK pathway and expression of *MMP9* in breast carcinoma cells [45]. The study showed that silencing of *EPCAM* inhibited *RAS* expression and phosphorylation level of *RAF*. The presence of *MMP9* expression suggested that the epithelial CTCs use *MMP9* to cleave the extra-cellular matrix so as they could be shed in the bloodstream.

In CTCs, the epithelial genes were expressed moderately and in low level in cluster 2 and cluster 3 when compared to cluster 1. EMT was represented in cluster 2 and mesenchymal/stem was represented in CTCs in cluster 3 (Fig. 5). EMT and stem cells markers, e.g. *CD133* and *CD44*, could be detected in epithelial CTCs, which were predictors of poor prognosis of patients with CRC [26]. In CTCs with EMT, EMT markers (*SNAI1* and *SLUG*) and one of the stem cells markers *CD133* were overexpressed, but low level of *MMP9* expression suggesting a complicated interplay among the EMT and stem cell markers which inhibited extra-cellular matrix degrading gene, *MMP9*. In this study, we found that *CD44* was not expressed in most of the CTCs. *CD44*, a receptor of the hyaluronic acid, has been reported to act as a docking molecule for *MMP9* [19]. This suggests that *MMP9* could not be compartmentalized in the CTCs due to the absence of *CD44*. At a certain time of their life cycle, CTCs do not need *MMP9* because they have already released from the primary tumour.

In mesenchymal/stem CTCs, the stemness markers (*CD133* and *CD44*) and *MMP9* were highly expressed, and epithelial genes were expressed in the lower level as expected. *CD133* and *CD44* can induce *MMP9* gene, but the expression of the genes *CD133* and *CD44* was stage-dependant. In concordance with the previous study, the current study has shown an overexpression of *MMP9* in *CD133*+/*CD44*+ cells than *CD133*+/*CD44*- cells [46]. Moreover, *CD133* was involved in the drug resistance mechanism of cancer [46]. We hypothesise that the dependency of the stemness protein was altered from *CD133* to *CD44* during the EMT to mesenchymal transition.

In this study, that presence of CTC was correlated with the biological aggressiveness of the CRC in terms of advanced clinical stages and adenocarcinoma of high grade. The findings are in concur with the theory that the evaluation of CTC is important in the monitoring of the progress of the disease after surgery, especially in the detection of cancer recurrence and metastases.

Conclusion

This study provides an improved negative selection technique to identify and characterise CTC subclones from patients with CRC. The findings of the study may help in the understanding of heterogeneities of CTC in patients with CRC and the roles of CTC in monitoring cancer progression. In addition, the study of phenotypic and genetic heterogeneities of CTC is important in research on CTC which could be used to determine the selection of drug selection and monitoring the therapy efficacy, drug-resistance and clonal evolution of cancers.

Author contributions

FBH and VG designed the project. FBH conducted the experiments, analysed the results and drafted the manuscript. C-TL and MM provided clinical information, patient samples and consents. TC assisted in sample collection and revised the manuscript. VG and AKL provided overall supervision and revision of the manuscript.

Acknowledgements

The authors would like to thank members of the Cancer Molecular Pathology Group and staff in Gold Coast University Hospital for supporting the collection of samples.

Funding

Griffith University and Gold Coast University Private Practice Trust Fund Grant of Gold Coast University Hospital.

Declarations of interest

The authors have no conflict of interest.

References

[1] Yokobori T, Iinuma H, Shimamura T, et al. Plastin3 is a novel marker for circulating tumor cells undergoing the epithelial-mesenchymal transition and is associated with colorectal cancer prognosis. Cancer Res. 2013; 73:2059-2069. https://doi.org/10.1158/0008-5472.CAN-12-0326

[2] Cohen SJ, Punt CJ, Iannotti N, et al. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. J Clin Oncol. 2008; 26:3213-3221. https://doi.org/10.1200/JCO.2007.15.8923

[3] Kin C, Kidess E, Poultsides GA, Visser BC, Jeffrey SS. Colorectal cancer diagnostics: biomarkers, cell-free DNA, circulating tumor cells and defining heterogeneous populations by single-cell analysis. Expert Rev Mol Diagn. 2013; 13:581-599. https://doi.org/10.1586/14737159.2013.811896

[4] Zhang J, Chen K, Fan ZH. Circulating Tumor Cell Isolation and Analysis. Adv Clin Chem. 2016; 75:1-31. https://doi.org/10.1016/bs.acc.2016.03.003

[5] Kulasinghe A, Perry C, Warkiani ME, et al. Short term ex-vivo expansion of circulating head and neck tumour cells. Oncotarget. 2016; 7:60101-60109. https://doi.org/10.18632/oncotarget.11159

[6] Liu Z, Fusi A, Klopocki E, et al. Negative enrichment by immunomagnetic nanobeads for unbiased characterization of circulating tumor cells from peripheral blood of cancer patients. J Transl Med. 2011; 9:70. https://doi.org/10.1186/1479-5876-9-70

[7] Tsai WS, Chen JS, Shao HJ, et al. Circulating Tumor Cell Count Correlates with Colorectal Neoplasm Progression and Is a Prognostic Marker for Distant Metastasis in Non-Metastatic Patients. Sci Rep. 2016; 6:24517. https://doi.org/10.1038/srep24517

[8] Wu W, Zhang Z, Gao XH, et al. Clinical significance of detecting circulating tumor cells in colorectal cancer using subtraction enrichment and immunostaining-fluorescence in situ hybridization (SE-iFISH). Oncotarget. 2017; 8:21639-21649. https://doi.org/10.18632/oncotarget.15452

[9] Chen JF, Ho H, Lichterman J, et al. Subclassification of prostate cancer circulating tumor cells by nuclear size reveals very small nuclear circulating tumor cells in patients with visceral metastases. Cancer. 2015; 121:3240-3251. https://doi.org/10.1002/cncr.29455

[10] Yao N, Jan YJ, Cheng S, et al. Structure and function analysis in circulating tumor cells: using nanotechnology to study nuclear size in prostate cancer. Am J Clin Exp Urol. 2018; 6:43-54.

[11] Gill DM, Agarwal N, Hahn AW, et al. Impact of circulating tumor cell (CTC) nucleus size on outcomes with abiraterone acetate (AA) therapy in men with metastatic castration-resistant prostate cancer (mCRPC). J. Clin. Oncol. 2017; 35:253-253. DOI: 10.1200/JCO.2017.35.6_suppl.253

[12] Ferraldeschi R., McDaniel A., Krupa R., et al. CK- and small nuclear size circulating tumor cell (CTCs) phenotypes in metastatic castration-resistant prostate cancer (mCRPC). J. Clin. Oncol. 2014; 32:209-209. DOI: 10.1200/jco.2014.32.4_suppl.209

[13] Cayrefourcq L, Mazard T, Joosse S, et al. Establishment and characterization of a cell line from human circulating colon cancer cells. Cancer Res. 2015; 75:892-901. https://doi.org/10.1158/0008-5472.CAN-14-2613

[14] Alix-Panabières C, Cayrefourcq L, Mazard T, Maudelonde T, Assenat E, Assou S. Molecular Portrait of Metastasis-Competent Circulating Tumor Cells in Colon Cancer Reveals the Crucial Role of Genes Regulating Energy Metabolism and DNA Repair. Clin Chem. 2017; 63:700-713. https://doi.org/10.1373/clinchem.2016.263582

[15] Vu T, Datta PK. Regulation of EMT in Colorectal Cancer: A Culprit in Metastasis. Cancers (Basel). 2017; 9:171. https://doi.org/10.3390/cancers9120171

[16] Kang Y, Massagué J. Epithelial-mesenchymal transitions: twist in development and metastasis. Cell. 2004; 118:277-279. https://doi.org/10.1016/j.cell.2004.07.011

[17] Qiao B, Gopalan V, Chen Z, Smith RA, Tao Q, Lam AK. Epithelial-mesenchymal transition and mesenchymal-epithelial transition are essential for the acquisition of stem cell properties in hTERT-immortalised oral epithelial cells. Biol Cell. 2012; 104:476-489. https://doi.org/10.1111/boc.201100077

[18] Lowes LE, Allan AL. Circulating Tumor Cells and Implications of the Epithelial-to-Mesenchymal Transition. Adv Clin Chem. 2018; 83:121-181. https://doi.org/10.1016/bs.acc.2017.10.004

[19] Dufour A, Zucker S, Sampson NS, Kuscu C, Cao J. Role of matrix metalloproteinase-9 dimers in cell migration: design of inhibitory peptides. J Biol Chem. 2010; 285:35944-35956. https://doi.org/10.1074/jbc.M109.091769

[20] Steinert G, Schölch S, Niemietz T, et al. Immune escape and survival mechanisms in circulating tumor cells of colorectal cancer. Cancer Res. 2014; 74:1694-1704. https://doi.org/10.1158/0008-5472.CAN-13-1885

[21] Sieuwerts AM, Mostert B, Bolt-de Vries J, et al. mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients. Clin Cancer Res. 2011; 17:3600-3618. https://doi.org/10.1158/1078-0432.CCR-11-0255

[22] Gopalan V, Ebrahimi F, Islam F, et al. Tumour suppressor properties of miR-15a and its regulatory effects on BCL2 and SOX2 proteins in colorectal carcinomas. Exp Cell Res. 2018; 370:245-253. https://doi.org/10.1016/j.yexcr.2018.06.025

[23] Germano G, Mauri G, Siravegna G, et al. Parallel Evaluation of Circulating Tumor DNA and Circulating Tumor Cells in Metastatic Colorectal Cancer. Clin Colorectal Cancer. 2018; 17:80-83. https://doi.org/10.1016/j.clcc.2017.10.017

[24] Ozkumur E, Shah AM, Ciciliano JC, et al. Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. Sci Transl Med. 2013; 5:179ra47. https://doi.org/10.1126/scitranslmed.3005616

[25] Yu M, Bardia A, Aceto N, et al. Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. Science. 2014; 345:216-220. https://doi.org/10.1126/science.1253533

[26] Ning Y, Zhang W, Hanna DL, et al. Clinical relevance of EMT and stem-like gene expression in circulating tumor cells of metastatic colorectal cancer patients. Pharmacogenomics J. 2018; 18:29-34. https://doi.org/10.1038/tpj.2016.62

[27] Wu F, Zhu J, Mao Y, Li X, Hu B, Zhang D. Associations between the Epithelial-Mesenchymal Transition Phenotypes of Circulating Tumor Cells and the Clinicopathological Features of Patients with Colorectal Cancer. Dis Markers. 2017; 2017:9474532. https://doi.org/10.1155/2017/9474532

[28] Zhao R, Cai Z, Li S, et al. Expression and clinical relevance of epithelial and mesenchymal markers in circulating tumor cells from colorectal cancer. Oncotarget. 2017; 8:9293-9302. https://doi.org/10.18632/oncotarget.14065

[29] Zhang D, Zhao L, Zhou P, et al. Circulating tumor microemboli (CTM) and vimentin+ circulating tumor cells (CTCs) detected by a size-based platform predict worse prognosis in advanced colorectal cancer patients during chemotherapy. Cancer Cell Int. 2017; 17:6. https://doi.org/10.1186/s12935-016-0373-7

[30] Coumans FA, van Dalum G, Beck M, Terstappen LW. Filter characteristics influencing circulating tumor cell enrichment from whole blood. PLoS One. 2013; 8:e61770. https://doi.org/10.1371/journal.pone.0061770

[31] Park S, Ang RR, Duffy SP, et al. Morphological differences between circulating tumor cells from prostate cancer patients and cultured prostate cancer cells. PLoS One. 2014; 9:e85264. https://doi.org/10.1371/journal.pone.0085264

[32] Wang C, Mu Z, Chervoneva I, et al. Longitudinally collected CTCs and CTC-clusters and clinical outcomes of metastatic breast cancer. Breast Cancer Res Treat. 2017; 161:83-94. https://doi.org/10.1007/s10549-016-4026-2

[33] Au SH, Storey BD, Moore JC, et al. Clusters of circulating tumor cells traverse capillary-sized vessels. Proc Natl Acad Sci U S A. 2016; 113:4947-4952. https://doi.org/10.1073/pnas.1524448113

[34] Meng S, Tripathy D, Frenkel EP, et al. Circulating tumor cells in patients with breast cancer dormancy. Clin Cancer Res. 2004; 10:8152-8162. https://doi.org/10.1158/1078-0432.CCR-04-1110

[35] Molnar B, Ladanyi A, Tanko L, Sréter L, Tulassay Z. Circulating tumor cell clusters in the peripheral blood of colorectal cancer patients. Clin Cancer Res. 2001; 7:4080-4085.

[36] Rejniak KA. Circulating Tumor Cells: When a Solid Tumor Meets a Fluid Microenvironment. Adv Exp Med Biol. 2016; 936:93-106. https://doi.org/10.1007/978-3-319-42023-3_5

[37] Tripathi K, Garg M. Mechanistic regulation of epithelial-to-mesenchymal transition through RAS signaling pathway and therapeutic implications in human cancer. J Cell Commun Signal. 2018; 12:513-527. https://doi.org/10.1007/s12079-017-0441-3

[38] Nomura A, Banerjee S, Chugh R, et al. CD133 initiates tumors, induces epithelial-mesenchymal transition and increases metastasis in pancreatic cancer. Oncotarget. 2015; 6:8313-8322. https://doi.org/10.18632/oncotarget.3228

[39] Miyazaki H, Takahashi RU, Prieto-Vila M, et al. CD44 exerts a functional role during EMT induction in cisplatin-resistant head and neck cancer cells. Oncotarget. 2018; 9:10029-10041. https://doi.org/10.18632/oncotarget.24252

[40] Fotiadou PP, Takahashi C, Rajabi HN, Ewen ME. Wild-type NRas and KRas perform distinct functions during transformation. Mol Cell Biol. 2007; 27:6742-6755. https://doi.org/10.1128/MCB.00234-07

[41] Chen C, Zhao S, Karnad A, Freeman JW. The biology and role of CD44 in cancer progression: therapeutic implications. J Hematol Oncol. 2018; 11:64. https://doi.org/10.1186/s13045-018-0605-5

[42] Dhar M, Lam JN, Walser T, Dubinett SM, Rettig MB, Di Carlo D. Functional profiling of circulating tumor cells with an integrated vortex capture and single-cell protease activity assay. Proc Natl Acad Sci U S A. 2018; 115:9986-9991. https://doi.org/10.1073/pnas.1803884115

[43] Grillet F, Bayet E, Villeronce O, et al. Circulating tumour cells from patients with colorectal cancer have cancer stem cell hallmarks in ex vivo culture. Gut. 2017; 66:1802-1810. https://doi.org/10.1136/gutjnl-2016-311447

[44] Hodgkinson CL, Morrow CJ, Li Y, et al. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. Nat Med. 2014; 20:897-903. https://doi.org/10.1038/nm.3600

[45] Gao J, Liu X, Yang F, Liu T, Yan Q, Yang X. By inhibiting Ras/Raf/ERK and MMP-9, knockdown of EpCAM inhibits breast cancer cell growth and metastasis. Oncotarget. 2015; 6:27187-27198. https://doi.org/10.18632/oncotarget.4551

[46] Hou Y, Zou Q, Ge R, Shen F, Wang Y. The critical role of CD133(+)CD44(+/high) tumor cells in hematogenous metastasis of liver cancers. Cell Res. 2012; 22:259-272. https://doi.org/10.1038/cr.2011.139

FIGURE LEGENDS

Fig. 1. Validation of negative selection by spiking method. The recovery capacity of the negative selection method was assessed by spiking a known number (0, 1, 10, 100) of cells in 5ml peripheral blood from healthy individuals. **(A)** A diagram of the procedure of spike-in experiments. **(B)** Comparison of recovery rates in different colon cancer cells (SW-48, SW-480 and HCT-116).

Fig. 2. CTC counts in patients with CRC. (A) Profiling of a 6-antibodies panel to detect CTCs from five patients with colorectal carcinoma. **(B)** Evaluation of EPCAM and CK18 expression in 3 colon cancer cell lines (SW-48, SW-480 and HCT-116). **(C)** Representative images of CTCs and the micro emboli (CTMs) isolated from the patients with CRC (40x magnification). **(D)** Comparison of CTC counts between patients with CRC and healthy donors. **(E)** Comparison of the number of CTCs and CTMs in patients with CRC.

Fig. 3. Morphological and phenotypical diversities in CTCs and CTMs. (A) Representative image of the measurement of cell length of CTCs. **(B)** Comparison of the cell-length between isolated CTCs and the cells in the CTMs. **(C)** Representative images of $EPCAM^{Pos}CK18^{Pos}$ and $EPCAM^{Pos}CK18^{Neg}$ subclones. **(D)** Comparison of the number of $EPCAM^{Pos}CK18^{Pos}$ and $EPCAM^{Pos}CK18^{Neg}$ subclones in patients with CRC.

Fig. 4. Molecular heterogeneities in the CTCs and tumour samples. Unsupervised clustering of 14 genes in **(A)** CTCs and **(B)** Tumours were represented using the Euclidean distance and the clustering method (Ward's minimum variance). The values were calculated from the log10 value of the relative quantification of each gene. Red colour indicates high expression and blue colour indicates low expression of the genes.

Fig. 5. Comparison of the gene expression in the CTCs and cancer tissues. Fourteen gene-set including epithelial genes, oncogenes, tumour suppressor, EMT genes, stemness and extracellular matrix (ECM) degrading genes were compared between CTCs and primary tumour tissues.

Fig. 6. Correlation with pathological stages of CRC. (A) Comparison of the number of **(A)** CTCs and **(B)** CTMs in patients with CRC of early stages (Ior II) and advanced stages (III or IV) . Comparison of the number of **(C)** $EPCAM^{Pos}CK18^{Pos}$ and **(D)** $EPCAM^{Pos}CK18^{Neg}$ subclones of CTCs in patients with early stages and advanced stages of CRC. Association of **(E)** cell length and **(F)** gene expression of stemness markers (CD44, CD133) of CTCs with pathological stages.

Supplementary Fig. 1. WBC (white blood cell) counts in healthy donors and patients with CRC. WBCs were counted by analysing the nucleated cells with no EPCAM or CK18 fluorescent signals from five ml of blood of healthy donors and patients with CRC.