

## RESEARCH ARTICLE

# Surface expression marker profile in colon cancer cell lines and sphere-derived cells suggests complexity in CD26<sup>+</sup> cancer stem cells subsets

Lorena Vázquez-Iglesias<sup>1</sup>, Leticia Barcia-Castro<sup>1</sup>, Marta Rodríguez-Quiroga<sup>1</sup>, María Páez de la Cadena<sup>1</sup>, Javier Rodríguez-Berrocal<sup>1</sup> and Oscar J. Cordero<sup>2,\*</sup>

## ABSTRACT

Taking advantage of eight established cell lines from colorectal cancer patients at different stages of the disease and the fact that all of them could form spheres, cell surface biomarkers of cancer stem cells and epithelial-mesenchymal transition were tested. The aim was to investigate cancer stem cells and metastatic stem cells in order to provide functional characterization of circulating tumor cells and promote the development of new anti-metastatic therapies. Our model showed an important heterogeneity in EpCAM, CD133, CD44, LGR5, CD26 and E-cadherin expression. We showed the presence of a subset of E-cadherin<sup>+</sup> (some cells being E-cadherin<sup>high</sup>) expressing CD26<sup>+</sup> (or CD26<sup>high</sup>) together with the well-known CSC markers LGR5 and EpCAM<sup>high</sup>, sometimes in the absence of CD44 or CD133. The already described CD26<sup>+</sup>/E-cadherin<sup>low</sup> or <sup>negative</sup> and CD26<sup>+</sup>/EpCAM<sup>-</sup>/CD133<sup>-</sup> subsets were also present. Cell division drastically affected the expression of all markers, in particular E-cadherin, so newborn cells resembled mesenchymal cells in surface staining. CD26 and/or dipeptidyl peptidase 4 inhibitors have already shown anti-metastatic effects in pre-clinical models, and the existence of these CD26<sup>+</sup> subsets may help further research against cancer metastasis.

**KEY WORDS:** Colorectal cancer, Cancer stem cells, Epithelial-mesenchymal transition, Biomarkers, CD26, EpCAM, LGR5, E-cadherin, CD133

## INTRODUCTION

Metastasis accounts for the vast majority of deaths due to cancer because even if the primary tumor has been perfectly removed by surgery, tumor cells can have disseminated and established themselves in distant locations (Oskarsson et al., 2014; Pantel et al., 2008; Liu et al., 2014a,b; Miranda-Lorenzo et al., 2014; Driessens et al., 2012). Cancer stem cells (CSCs) are the only tumor cell type with long-term self-renewal potential because of their microenvironmental niche (Oskarsson et al., 2014; Liu et al., 2014a, b; Miranda-Lorenzo et al., 2014), suggesting that metastatic stem

cells (MetSCs) with tumor-initiating capacity already exist in the primary tumor (Oskarsson et al., 2014; de Sousa e Melo et al., 2017). Since metastasis does not rely on driver mutations, and genomic biomarkers are not useful for diagnosis (Vanharanta and Massagué, 2013), the possibility that MetSCs may be tracked is especially important for metastasis diagnosis and development of therapeutic approaches that kill them (Oskarsson et al., 2014).

Cell surface markers are powerful tools, for example, for isolating distinct cell populations from freshly harvested primary tumors (Miranda-Lorenzo et al., 2014) or blood. However, the expression levels of markers of CSCs change depending on environmental conditions; they are not reproducible across or even within similar tumor types, and they are not exclusive of a functional CSC phenotype (Miranda-Lorenzo et al., 2014; Driessens et al., 2012; de Sousa e Melo et al., 2017; Vanharanta and Massagué, 2013; Wicha et al., 2006). Thus, an important question is which markers should be used for CSC and MetSC characterization.

In colorectal cancer (CRC), MetSCs are already present in the primary tumor (de Sousa e Melo et al., 2017; Varela-Calviño and Cordero, 2015). Candidate markers for CRC MetSC characterization include CD166, CD29, CD24, LGR5, EpCAM (CD326), ALDH1, CDCP1, CXCR4, CC188 (Hsu et al., 2013; Pitule et al., 2014) and ephrin type B receptor 2 (EphB2) (Rowehl et al., 2014), although many of these markers are also expressed in normal colonic stem cells (i.e. LGR5, ALDH1, or CD29), complicating the distinction between CSCs and normal stem cells. From present knowledge, CRC MetSCs can be found among the cell population with a high expression of Wnt targets LGR5<sup>+</sup> and EphB2<sup>+</sup> (de Sousa e Melo et al., 2017; Jung et al., 2011; Kemper et al., 2012) co-expressing EpCAM, CD133, CXCR4 and CD26 markers. EpCAM, CD133 and CXCR4 are enriched in a metastatic cell population with an auto fluorescent subcellular compartment (Miranda-Lorenzo et al., 2014), and a CD26<sup>+</sup>/CD133<sup>+</sup>/CD44<sup>+</sup> CSC population was capable of metastasizing when transplanted in mice (Pang et al., 2010). Most of these markers are co-expressed in the primary tumor; so, it is expected that a particular marker combination may be used to identify MetSCs in CRC. Intriguingly, a CD26<sup>+</sup> circulating tumor cell (CTC) population that is CD44<sup>+</sup> and CD66c<sup>+</sup> but EpCAM<sup>-</sup> and CD133<sup>-</sup> is an independent prognostic factor for CRC recurrence (Lieto et al., 2015).

As the number of cells presenting CSC features that can be obtained from surgical samples is scarce (Oskarsson et al., 2014; de Sousa e Melo et al., 2017; Varela-Calviño and Cordero, 2015; James et al., 2015) and CTCs from liquid biopsies are highly heterogeneous (Khoja et al., 2015; Acosta et al., 2016), several studies have previously explored the possibility of using established cell lines. In spite of their inherent genetic instability during long-term passaging, CTCs are a valid option (Rowehl et al., 2014; Dotse and Bian, 2016), as they are able to sustain different cell subsets

<sup>1</sup>Department of Biochemistry, Genetics and Immunology, Facultade de Bioloxía, Universidade de Vigo, 36200 Vigo, Galicia, Spain (EU). <sup>2</sup>Department of Biochemistry and Molecular Biology. CIBUS Building, Facultade de Bioloxía. Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Galicia, Spain (EU).

\*Author for correspondence (oscarj.cordero@usc.es)

 O.J.C., 0000-0003-1026-124X

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

with CSC features (Zimmerer et al., 2013) and with the inherent heterogeneity observed in cancer populations (Driessens et al., 2012; Pitule et al., 2014; Lieto et al., 2015; Zimmerer et al., 2013). The aim of this study was to test candidate CSC and MetSC markers in a panel of eight cell lines obtained from primary or metastatic tumors of patients in different disease stages [so that epithelial-mesenchymal transition (EMT) markers were also used]. The results would provide a functional characterization of specific cell subsets found in clinical specimens (Lierto et al., 2015; Cheung et al., 2017) and the possibility of using these chemo-resistant cells as targets for the development of new therapies.

## RESULTS

### Expression of stem cell, CSC and EMT markers in CRC cell lines

The stem cell, CSC and EMT marker expression profile in eight CRC cell lines was analyzed by flow cytometry (Table 1), western blotting (Fig. 1A) and immunofluorescence (Fig. 1B–D). Cytometry results (Table 1) showed that cell lines SW1116, SW480 and SW620 presented the lowest E-cadherin expression. In western blotting (Fig. 1A), the mature form of E-cadherin (120 kDa) was detectable in DLD-1, HT-29, Caco-2, COLO205 and T84 but not in SW1116, SW480 and SW620, supporting the flow cytometry results. An E-cadherin high MW band (130 kDa), possibly corresponding to the inactive precursor (Khoja et al., 2015), was detected in all tested cell lines, although the expression was very low in SW480 and SW620. On the contrary, vimentin was only expressed in SW480 and SW620 (Fig. 1A), as expected for a mesenchymal stage and according to the expression observed for E-cadherin.

A very high frequency of EpCAM<sup>+</sup> cells was found in all cell lines (Table 1) and a band of 40 kDa was observed upon western blotting for all cell lines (Fig. 1A), suggesting that these cell lines were fully differentiated carcinoma cells. On the contrary, all cell lines showed a percentage of stem cells from intestinal origin as indicated by LGR5 expression, with the highest frequency in T84 cells (22.5%).

Cell lines with the lowest E-cadherin-expression showed low frequencies of CD44 (SW1116 and SW620), CD26 (SW480 and SW620), LGR5 (SW620), and CD133 (SW1116 and SW480) (Table 1) by flow cytometry. However, the expression of these CSC markers was very heterogeneous in the other cell lines, CD133 and CD44 being the most variable (between 1% in T84 and 82.9% in Caco-2 and between 10.1% in COLO205 and 85.9% in Caco-2, respectively). In the case of CD26, all the cell lines showed a high expression, except SW480, SW620 and DLD-1 (intermediate frequency of positive cells). In DLD-1 and COLO205, there was a lack of correlation between western blotting (Fig. 1A) and flow cytometry data (Table 1 and Fig. S1), which should be further investigated.

We also analyzed the autofluorescence phenotype of cell lines (exclusive of epithelial CSC), which could be enhanced using riboflavin (Miranda-Lorenzo et al., 2014). All the analyzed cell lines had a very small subpopulation of CSCs (Table S2).

Immunofluorescence (Fig. 1B–D) showed that the E-cadherin and EpCAM staining distributions were very similar and corresponded to those of proteins related to cell–cell interaction. The distributions of CD44, CD26 and LGR5 staining were more diffused, although they were plasma membrane proteins. However, CD26 staining showed a trend for polarization that LGR5 lacked.

### Phenotypic characterization of subsets in the cell lines

Current knowledge of normal and tumor tissues indicates that CSCs are rarely defined by a single marker but by a combination of multiple molecular markers. On the other hand, several studies have linked high surface expression of some of these markers with the tumor degree of differentiation, depth of invasion, clinical stage and metastatic status in CRC (Acosta et al., 2016; Dotse and Bian, 2016; Ren et al., 2013; LaBarge and Bissell, 2008; Kojima et al., 2008; Horst et al., 2008). Therefore, we established all possible combinations between the markers and all possible combinations of subpopulations with high intensity expression of markers.

For double positive subsets (Table S3), we observed high frequencies of almost all the markers in HT-29 and Caco-2 and very low frequencies in SW480 and SW620, as expected. The other cell lines showed different expressions depending on the analyzed subset. We observed that almost all LGR5<sup>+</sup> cells in the cell lines were EpCAM<sup>+</sup> (Tables 1 and S3) and had higher frequency of CD26<sup>+</sup>/E-cadherin<sup>+</sup> cells than of CD133<sup>+</sup>/E-cadherin<sup>+</sup> cells.

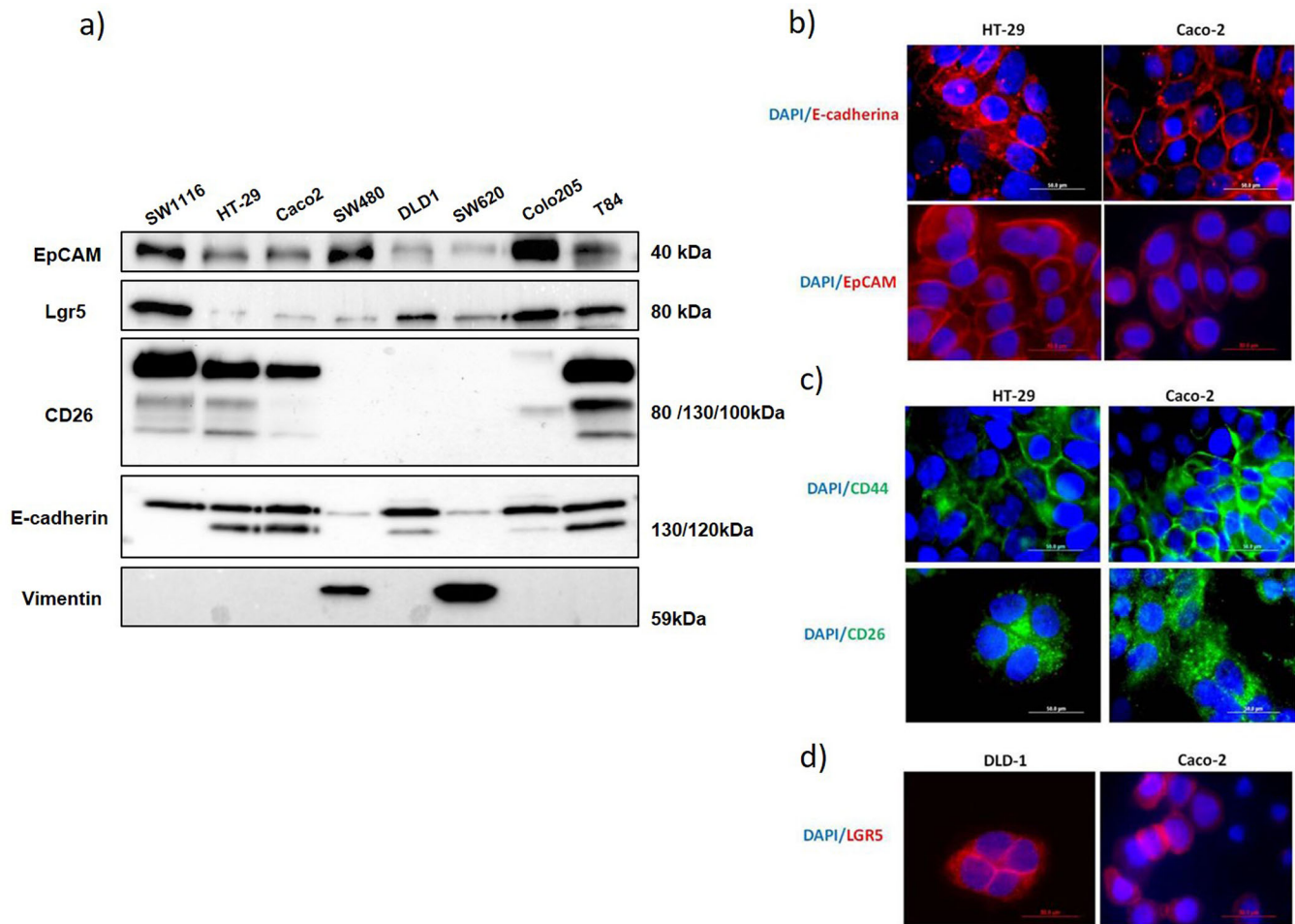
Markers with high expression in the cell lines were also observed. Fig. S2 shows dot plots for LGR5 versus EpCAM and CD133 versus CD26 in COLO205 and Caco cell lines, as examples of the gating strategy. Cells with high expression of EpCAM were easily detected in COLO205 and Caco cell lines, and all the cell lines had EpCAM<sup>high</sup>/LGR5<sup>+</sup> subsets (from 2.8% in SW620 to 11.9% in Caco-2, Table 2), although not all LGR5<sup>+</sup> cells were EpCAM<sup>high</sup>. In the case of dot plots for CD133 versus CD26 (not studied in this context before) (Fig. S2B, Table 3), one cell line (Caco-2) showed large CD133<sup>+</sup>CD26<sup>high</sup> and CD133<sup>high</sup>/CD26<sup>high</sup> subsets (65.3% and 9.6%, Fig. S2B). Very small CD133/CD26 subsets with high expression of one or both markers were observed in seven out of the eight analyzed cell lines (Table 3). In addition, the cell lines had CD133<sup>-</sup>/CD26<sup>high</sup> subpopulations, except the mesenchymal cell lines (Table S4).

We also looked at possible E-cadherin<sup>high</sup> subsets. Most cell lines (SW1116, DLD-1, HT-29, Caco-2, COLO205 and T84) harbored CD26<sup>high</sup>/E-cadherin<sup>high</sup>, CD26<sup>high</sup>/E-cadherin<sup>+</sup>, or E-cadherin<sup>high</sup>/CD26<sup>+</sup> subsets (Table S5), but SW480 and SW620 did not show any of these subsets. CD133<sup>high</sup> associated to E-cadherin<sup>high</sup>

**Table 1. Frequencies of CD26, CD133, CD44, EpCAM, LGR5 and E-cadherin in colon cancer cell lines**

Cell line	CD26 <sup>+</sup> (%) M±s.d.	CD133 <sup>+</sup> (%) M±s.d.	CD44 <sup>+</sup> (%) M±s.d.	EpCAM <sup>+</sup> (%) M±s.d.	LGR5 <sup>+</sup> (%) M±s.d.	E-cadherin <sup>+</sup> (%) M±s.d.
SW1116	81.1±10.1	7.7±1.2	23.4±4.0	91.4±5.8	10.2±1	43.0±4.5
SW480	2.7±0.8	3.4±2.1	37.0±1.0	91.8±8.8	8.0±1.8	43.4±16.8
DLD-1	41.9±9.1	16.7±3.4	58.3±12.0	85.2±15.2	12.8±2.7	82.6±11.3
SW620	2.3±0.6	60.3±22.8	6.1±1.8	90.0±11.2	3.7±1.3	30.8±6.0
HT-29	75.1±6.8	73.9±6.6	78.1±13.4	97.6±2.7	9.1±4.1	83.9±8.7
Caco-2	86.8±12.1	82.9±10.7	85.9±9.8	91.1±8.3	15.9±6.4	83.2±14.2
COLO205	76.0±9.9	81.8±7.6	10.1±0.3	99.1±1.7	9.9±1.3	90.4±11.4
T84	80.7±4.8	1.0±0.1	25.3±19.0	87.1±1.9	22.5±4.4	76.6±14.6

M: Mean of three experiments.  
s.d.: Standard deviation.



**Fig. 1. Expression of different markers in the eight human colon cancer cell lines analyzed.** (A) Western blot analysis of EpCAM, LGR5, CD26, E-cadherin and vimentin expression in total cell extracts from the eight cell lines (20 μg of protein in each line). Data shown are representative of three experiments. (B) E-cadherin and EpCAM expression analysis by immunofluorescence in HT-29 and Caco-2 cells. (C) CD44 and CD26 expression analysis by immunofluorescence in HT-29 and Caco-2 cells. (D) LGR5 expression analysis by immunofluorescence in DLD-1 and Caco-2 cells. Nuclei were stained with DAPI. Scale bars: 50 μm.

was practically non-existent, and CD133 staining was poor in E-cadherin<sup>+</sup> cells (Table 1 and Table S6).

No subset with high expression of CD44 was observed (data not shown), but the cells from the CD133<sup>high</sup>/CD26<sup>high</sup> subset in the Caco-2 cell line were CD44<sup>-</sup> and E-cadherin<sup>low</sup>.

#### Sphere formation in the CRC cell lines under investigation

Whether the cell lines which originated from both primary and metastatic tumors were able to form spheres was examined. All cell lines formed spheres in the first generation after 7 days of culture (Fig. S3), suggesting the presence of CSCs in all cell lines. However, there were appreciable differences in the structure and size of the spheres (Fig. S3 and Table S7). SW480, SW620 and COLO205 spheres showed similar morphology and were different from the other cell line spheres. Those three cell lines had a low expression of E-cadherin (120 kDa) (Fig. 1B) as well as a low frequency of CD44 (Table 1). A lack or low expression of E-cadherin and CD44 could lead to decreased cell-to-cell contacts and to the observed morphology. However, SW1116 cell line also showed similar marker expression.

The sphere cells were disaggregated, and cells reseeded to establish self-renewal capacity by formation of secondary spheres

(Table S7). Only SW1116 did not form secondary spheres. All the other cell lines did form spheres in three serial passages: DLD-1 formed more spheres and T84 less spheres (data not shown). Efficiency for self-renewal was essentially maintained through passages in all cell lines, T84 cells being the most efficient and SW620 the least efficient (Table S7).

#### Expression of stem cell, CSC and EMT markers in cells derived from spheres

The spheres developed in 7-day cultures were disaggregated, and stem cell, CSC and EMT markers were analyzed by flow cytometry in sphere-derived cells (<sub>sph</sub>) (Table 4).

Sphere-derived cells from all cell lines showed high frequencies of E-cadherin expression, except SW620 (Table 4). As expected for proliferating cells in epithelial state, E-cadherin frequencies were much higher in SW1116<sub>sph</sub> and SW480<sub>sph</sub> than those in their respective cell lines (Table 1). However, this was not the case in SW620<sub>sph</sub> and T84<sub>sph</sub>, where the frequencies of E-cadherin<sup>+</sup> were similar to the original cell line (being a minority in SW620).

More than half of SW1116<sub>sph</sub> (62%) was LGR5<sup>+</sup>, with approximately 20% of positivity for this marker in the rest of cell lines<sub>sph</sub>, except SW620<sub>sph</sub>, where only a minor subset (2.8%) was

**Table 2. Frequencies of EpCAM<sup>high</sup>/LGR5<sup>+</sup> and EpCAM<sup>low</sup>/LGR5<sup>+</sup> subsets in colon cancer cell lines**

Cell line	EpCAM <sup>high</sup> /LGR5 <sup>+</sup> % M±s.d.	EpCAM <sup>low</sup> /LGR5 <sup>+</sup> % M±s.d.
SW1116	6.8±3.1	2.9±3.1
SW480	2.9±2.3	2.0±0.5
DLD-1	6.8±1.1	1.5±0.3
SW620	2.8±1.3	0.5±0.3
HT-29	7.4±3.4	1.7±1.6
Caco-2	11.9±8.5	3.3±3.2
COLO205	6.6±4.4	1.6±1.9
T84	8.4±4.2	9.8±6.4

M: Mean of three experiments.  
s.d.: Standard deviation.

LGR5<sup>+</sup> (Table 4). Sphere-derived cells showed enhanced frequencies (two- to three-fold) of LGR5 positivity compared to the original cell lines (Tables 4 and 1). All the cell lines<sub>sph</sub> showed high frequencies of EpCAM<sup>+</sup> (Table 4). SW1116<sub>sph</sub>, with high frequencies of LGR5<sup>+</sup>, also showed a high frequency (around a 50%) of the LGR5<sup>+</sup>/EpCAM<sup>high</sup> subset (Table S8, Fig. 2, region B of dot plots). On the contrary, more EpCAM<sup>low</sup> were found in SW480<sub>sph</sub>, SW620<sub>sph</sub> and COLO205<sub>sph</sub> (SW480<sub>sph</sub> and SW620<sub>sph</sub> showing the lowest frequencies of LGR5<sup>+</sup>). In addition, the large subset of LGR5<sup>+</sup>/EpCAM<sup>high</sup> in T84<sub>sph</sub> showed some LGR5<sup>high</sup> cells (Fig. 2, red arrow), which has not been described before.

The frequencies of CD133, CD44 and CD26 in cell lines<sub>sph</sub> resembled those of the original lines, although there were some changes: CD133<sup>+</sup> frequencies in T84<sub>sph</sub> were higher than those in the original cell line (17%), whereas HT-29<sub>sph</sub> and COLO205<sub>sph</sub> showed lower CD133 frequencies than those in the lines of origin (Fig. S4). CD44 was also found at high frequencies in cell lines<sub>sph</sub>, except in SW620<sub>sph</sub> (Table 4). Frequencies were also low in SW480<sub>sph</sub> compared to the sphere-derived cells from the other cell lines. CD26 showed particularly higher frequencies in SW480<sub>sph</sub> than those in the original cell line, despite being strikingly similar to the original cell lines in the other cases (Tables 1 and 4).

Very interestingly, CD133, CD44 and CD26 hardly correlated (Fig. S4). Intriguingly, although the same cell lines continued to show a CD26<sup>high</sup> population in sphere-derived cells, the frequency was quite a lot lower in Caco-2<sub>sph</sub> and much higher in T84<sub>sph</sub> than those in the original cell lines (Tables S9 and S4). To note, the latter are the ones with EpCAM<sup>high</sup>/LGR5<sup>high</sup>. The CD133<sup>high</sup>/CD26<sup>high</sup> subsets were very small in cell lines<sub>sph</sub> but more detectable than in the original cell lines (Table S10). However, the subsets of CD133<sup>high</sup>/CD26<sup>-</sup> cells in some cell lines (e.g. SW620<sub>sph</sub> with

4.2%) or in the other combinations (Table S10) showed higher frequencies than those in the original cell lines.

### Relationship between CSC and EMT markers in cells derived from spheres

All cell lines<sub>sph</sub> showed some E-cadherin<sup>-</sup> cells and E-cadherin<sup>+</sup> subsets corresponding to mesenchymal and epithelial cells, respectively. The subsets of E-cadherin<sup>-</sup> cells in all cell lines<sub>sph</sub> had a higher percentage of small size cells than that of the subsets of E-cadherin<sup>+</sup> cells (e.g. DLD-1<sub>sph</sub> in Fig. 5). We analyzed if CSC markers were differentially expressed in both E-cadherin<sup>-</sup> and E-cadherin<sup>+</sup> sphere-derived cells.

### Markers in E-cadherin<sup>-</sup> sphere-derived cells

Most but not all E-cadherin<sup>-</sup> cells are also CD133<sup>-</sup>, CD26<sup>-</sup>, or CD44<sup>-</sup> (e.g. CD133 in Fig. 5). A common characteristic in these E-cadherin<sup>-</sup> cells was the presence of one or more subsets expressing only one marker in different combinations in all cell lines<sub>sph</sub>, and co-expression of CD133/CD26 was not found in five of them. SW1116<sub>sph</sub> was CD133<sup>+</sup>, CD26<sup>+</sup> and CD44<sup>+</sup>, and had a few CD26<sup>+</sup>/CD44<sup>+</sup> cells (2.3%); SW480<sub>sph</sub> and SW620<sub>sph</sub> were totally CD26<sup>-</sup>, with a high percentage of CD133<sup>+</sup> and a low percentage of CD44<sup>+</sup> or CD133<sup>+</sup>/CD44<sup>+</sup>; DLD-1<sub>sph</sub> had small subsets of only CD26<sup>+</sup>, CD44<sup>+</sup> and CD133<sup>+</sup> cells; HT-29<sub>sph</sub> had a small subset of CD133<sup>+</sup>/CD26<sup>+</sup>/CD44<sup>+</sup>, CD44<sup>+</sup>/CD26<sup>+</sup>, or only CD44<sup>+</sup> cells; Caco-2<sub>sph</sub> had a large CD133<sup>+</sup>/CD26<sup>+</sup>/CD44<sup>+</sup> and CD26<sup>+</sup>/CD44<sup>+</sup> subsets and a small CD26<sup>+</sup> or CD44<sup>+</sup> subset; COLO205<sub>sph</sub> had a large CD26<sup>+</sup> subset and a small CD133<sup>+</sup>/CD26<sup>+</sup> subset; T84<sub>sph</sub> was mostly CD133<sup>-</sup>, with a high CD26<sup>+</sup> percentage and a low percentage of CD44<sup>+</sup> or CD26<sup>+</sup>/CD44<sup>+</sup>.

### Markers in E-cadherin<sup>+</sup> sphere-derived cells

Although still having small cells, the majority of the cells in the E-cadherin<sup>+</sup> subset were large, and almost all large cells were in this subset (Fig. S5). The frequency of the CD26<sup>+</sup>/CD44<sup>+</sup> subset increased independently of the expression of CD133, although there were still cells only CD26<sup>+</sup> or CD44<sup>+</sup> (data similar to Fig. S4). Moreover, this CD26<sup>+</sup>/CD44<sup>+</sup> subset correlated with the stage of origin of the cell line (larger subsets in cell lines from advanced stage tumors, data not shown), irrespective of the presence of CD133.

A correlation between E-cadherin and CD26 was observed in most cell lines<sub>sph</sub>. A population of E-cadherin<sup>high</sup> cells was observed in all cell lines<sub>sph</sub> (Fig. S6, regions A and B), except in SW620<sub>sph</sub>, and these cells were mainly CD26<sup>high</sup> (Fig. S6, region B, and Table S11). A similar profile was found many times for CD44 (data not shown). However, in the case of CD133, although some E-cadherin<sup>high</sup> cells showed CD133 staining (Table S6), most CD133<sup>+</sup> cells were E-cadherin<sup>-</sup> (Table 4 and Table S3).

Interestingly, positivity for E-cadherin and LGR5 and their fluorescence intensity were strongly correlated (Fig. 3), and no LGR5<sup>+</sup> cells were E-cadherin<sup>-</sup>. These cells were also more complex in morphology (SSC axes in Fig. 3B). As shown in Fig. 3B (region 6 in the UR quadrant) and from other results not shown, we confirmed that when CD26<sup>high</sup> cells were present, they were LGR5<sup>+</sup>. However, there were some LGR5<sup>+</sup> cells without CD26 and CD133 markers (e.g. Caco-2<sub>sph</sub> in Fig. 3C).

Therefore, as LGR5 and E-cadherin expression correlate in sphere cells (Fig. 3A) – as well as LGR5 with EpCAM (Fig. 2) and with CD26 (Fig. 3B), and also E-cadherin with CD26 (Fig. S6) but not with CD133 (Fig. 3C) – it can be deduced that, in sphere derived cells from lines of advanced tumor stage, the E-cadherin<sup>+</sup>/LGR5<sup>+</sup>/EpCAM<sup>high</sup> cells are also CD26<sup>high</sup>. Some of these E-cadherin<sup>+</sup>

**Table 3. Frequencies of CD26/CD133 subsets in colon cancer cell lines**

Cell line	CD26 <sup>high</sup>		CD133 <sup>high</sup> CD26 <sup>low</sup> % M±s.d.
	CD133 <sup>high</sup> % M±s.d.	CD133 <sup>low</sup> % M±s.d.	
SW1116	0.08±0.15	3.05±2.15	0.03±0.04
SW480	-	-	-
DLD-1	-	0.03±0.05	0.05±0.10
SW620	-	-	0.40±0.55
HT-29	0.30±0.18	1.13±0.21	0.20±0.08
Caco-2	9.58±11.34	12.13±11.98	0.38±0.22
COLO205	-	0.05±0.06	0.15±0.10
T84	-	1.33±1.03	-

M: Mean of three experiments.  
s.d.: Standard deviation.  
-: Not detected.

**Table 4. Frequencies of CD26, CD133, CD44, EpCAM, LGR5 and E-cadherin in cells disaggregated from spheres grown from colon cancer cell lines**

Cell line <sub>sph</sub>	CD26 <sup>+</sup> % M±s.d.	CD133 <sup>+</sup> % M±s.d.	CD44 <sup>+</sup> % M±s.d.	EpCAM <sup>+</sup> % M±s.d.	LGR5 <sup>+</sup> % M±s.d.	E-cadherin <sup>+</sup> % M±s.d.
SW1116	84.6±12.9	30.2±19.7	52.2±12.9	84.4±8.7	62.1±12.3	84.4±6.3
SW480	22.1±10.3	31.3±26.7	35.6±20.2	82.7±12.6	22.3±3.3	73.8±12.1
DLD-1	47.7±28.2	24.5±9.3	77.5±6.9	84.8±11.2	26.1±21.3	86.5±8.8
SW620	1.8±0.7	62.6±14.5	6.1±5.7	91.6±7.6	2.8±2.7	35.4±26.1
HT-29	92.4±8.6	54.2±11.5	77.6±12	93.6±4.9	22.3±27.9	84.5±13.5
Caco-2	98.0±2.9	77.8±4.8	96.8±2.9	97.5±3	16.9±19	90.0±8.8
COLO205	85.7±1.1	73.0±17.4	85.0±1.7	77.3±18.1	19.5±11.9	86.8±14.1
T84	73.6±8.3	17.6±6.7	70.7±6.5	79.0±11.5	29.0±7.6	79.1±16.4

Cell line<sub>sph</sub>: Spheres derived-cells for each cell line.

M: Mean of three experiments.

s.d.: Standard deviation.

LGR5<sup>+</sup>/EpCAM<sup>high</sup>/CD26<sup>high</sup>, but not all, were CD133<sup>+</sup> or CD133<sup>high</sup> (Fig. 3C), and most were CD44<sup>+</sup> (Fig. S4).

### Clustering of markers in sphere-derived cells

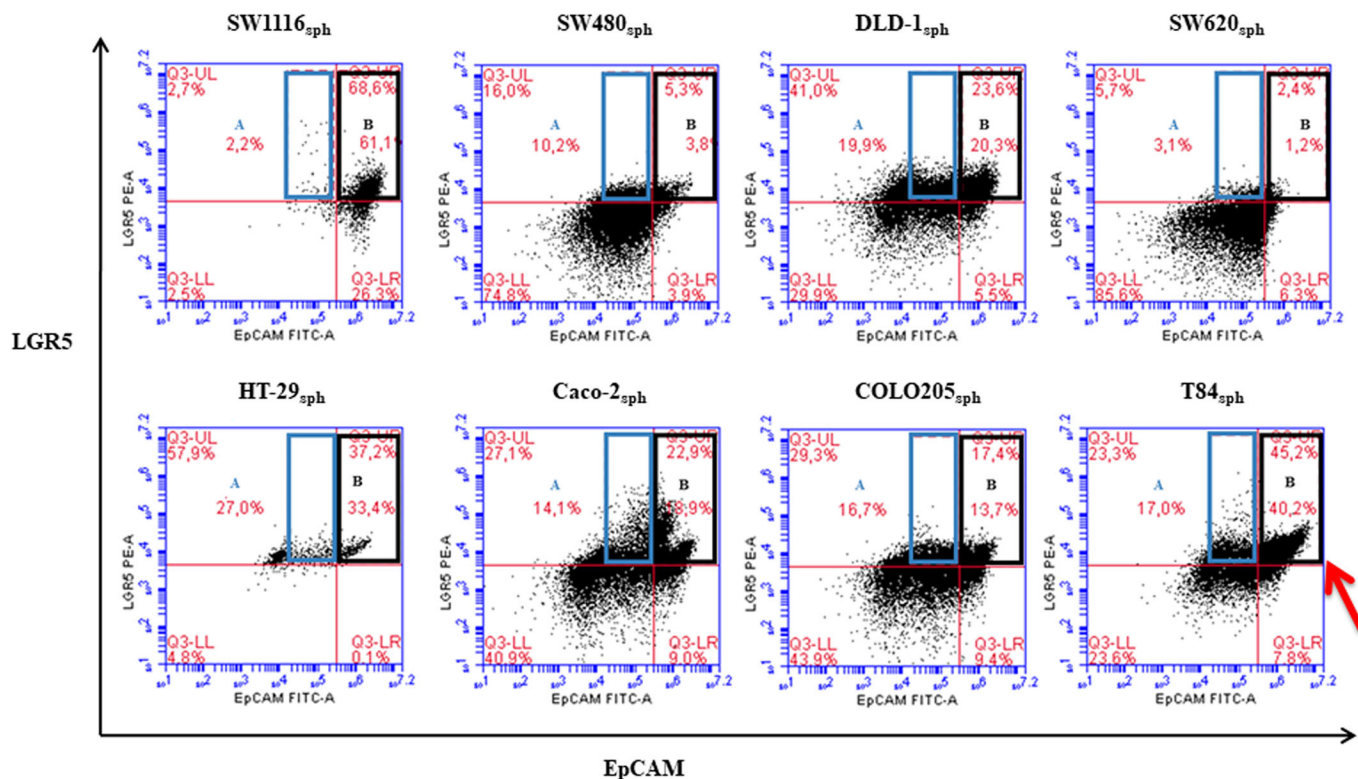
To confirm the above finding, LGR5<sup>+</sup>/EpCAM<sup>high</sup> subsets from T84<sub>sph</sub> were sorted into E-cadherin<sup>low</sup> and E-cadherin<sup>high</sup> subsets, which were analyzed for CD26, CD44 and CD133 expression (Fig. 4). Both subsets showed low expression of CD133 and CD44. The E-cadherin<sup>high</sup> subset was mostly CD26<sup>+</sup> (82%), while the E-cadherin<sup>low</sup> subset was only 20% CD26<sup>+</sup>.

### DISCUSSION

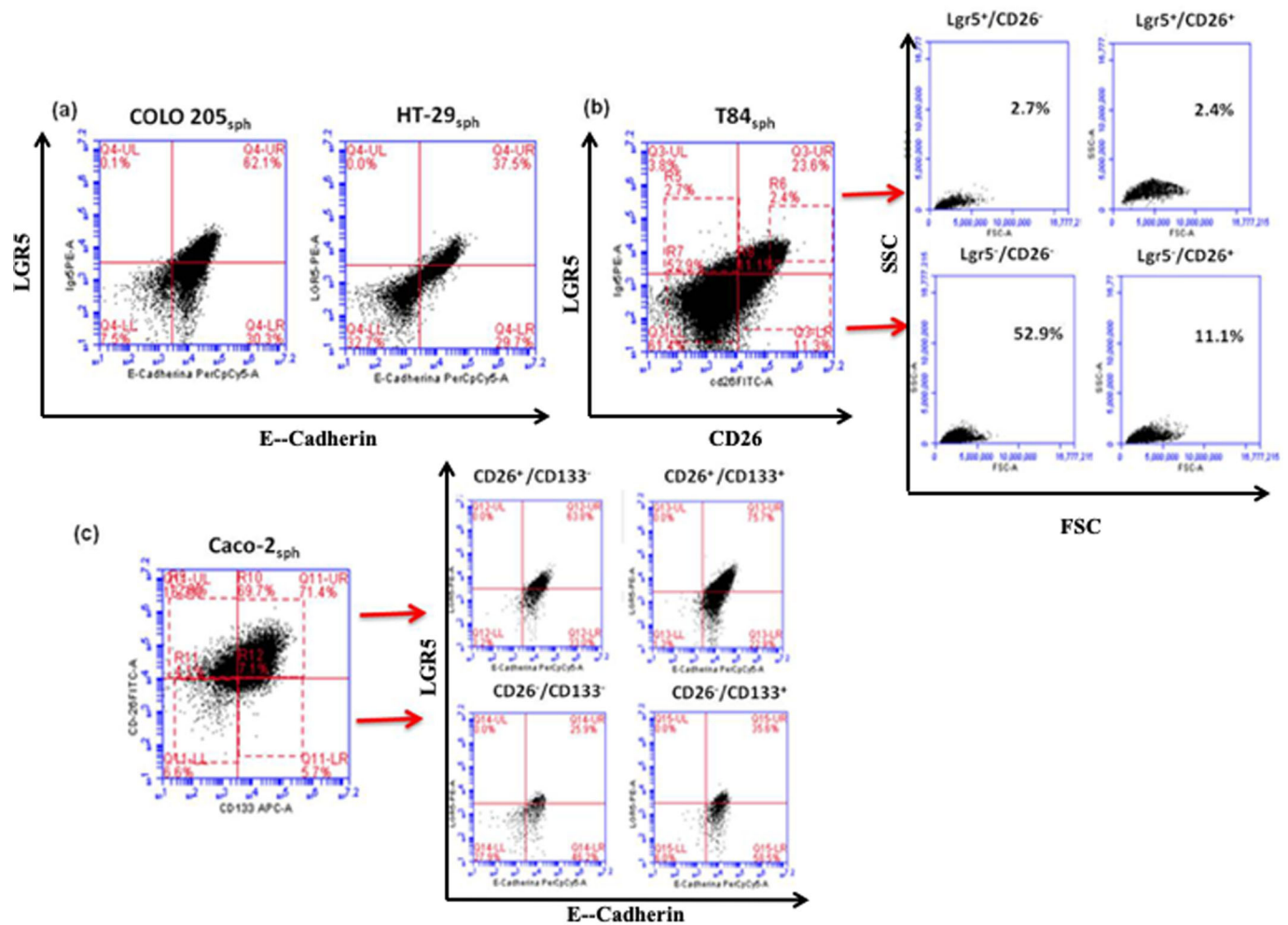
We show for the first time that LGR5<sup>+</sup>, E-cadherin<sup>high</sup>, EpCAM<sup>high</sup> and CD26<sup>high</sup> are frequently associated in sphere-derived cells, that CD133 seems to be related to a different germinal line, and that cell division affects the expression of all markers, including that of E-cadherin. These results are highlighted with the recent report that has shown that LGR5<sup>+</sup> cells are more important for the process of

metastasis than for primary tumor growth (de Sousa e Melo et al., 2017).

Circulating human CD133<sup>+</sup>/CD26<sup>+</sup>/CD44<sup>+</sup> cells but not CD133<sup>+</sup>/CD26<sup>-</sup>/CD44<sup>+</sup> cells have been detected in the portal vein of mice at week 6 after cecal wall injection, demonstrating the invasion of CD26<sup>+</sup> cells into the circulation of orthotopically implanted mice, leading to the development of liver metastasis (Pang et al., 2010). Interestingly, E-cadherin expression was down-regulated in CD133<sup>+</sup>/CD26<sup>+</sup> cells from primary CRC tumors compared to CD133<sup>+</sup>/CD26<sup>-</sup> cells, with the concomitant upregulation of N-cadherin, the E-cadherin repressor slug, as well as other EMT markers, such as twist, fibronectin and vimentin (Pang et al., 2010; Cheung et al., 2017). Here, we show that, in most analyzed cell lines, CD133 and CD26 are hardly expressed together and CD133<sup>-</sup>/CD26<sup>+</sup> sphere cells (perhaps CSCs) are a major population. Although the subset E-cadherin<sup>low</sup> or<sup>-</sup>/CD133<sup>+</sup>/CD26<sup>+</sup> could be found in all lines and also in sphere-derived cells, the E-cadherin<sup>high</sup>/CD133<sup>-</sup>/CD26<sup>+</sup> subset was particularly large in cell



**Fig. 2. Flow cytometry analysis of LGR5/EpCAM in sphere-derived cells.** T84<sub>sph</sub> showed a subset with LGR5 high expression (LGR5<sup>high</sup>, red arrow). LGR5<sup>+</sup>/EpCAM<sup>low</sup> region A, LGR5<sup>+</sup>/EpCAM<sup>high</sup> region B.



**Fig. 3.** Flow cytometry analysis of LGR5/E-cadherin, LGR5/CD26 and CD26/CD133 in sphere-derived cells. (A) Two representative dot plots of LGR5 versus E-cadherin expression in COLO205<sub>sph</sub> and HT-29<sub>sph</sub>. (B) LGR5 versus CD26 expression in T84<sub>sph</sub>. CD26<sup>high</sup>/LGR5<sup>+</sup> cells are marked in R6 region in UR quadrant. Physical gatings (FCS versus SSC) of the four quadrants of T84<sub>sph</sub> are shown. (C) Caco-2<sub>sph</sub> representative dot plots for CD26 versus CD133 and, on the right, LGR5 versus E-cadherin dot plots of the four regions gated on the left CD26/CD133 dot plot.

lines from advanced CRC stages. This result is consistent with the fact that in humans, the population isolated from the primary tumor comprising CD133<sup>+</sup>/CD44<sup>+</sup>/CD26<sup>+</sup> cells (and E-cadherin<sup>low</sup> or E-cadherin<sup>-</sup>) is not the only CSC population present in the tumor biopsies (Pang et al., 2010).

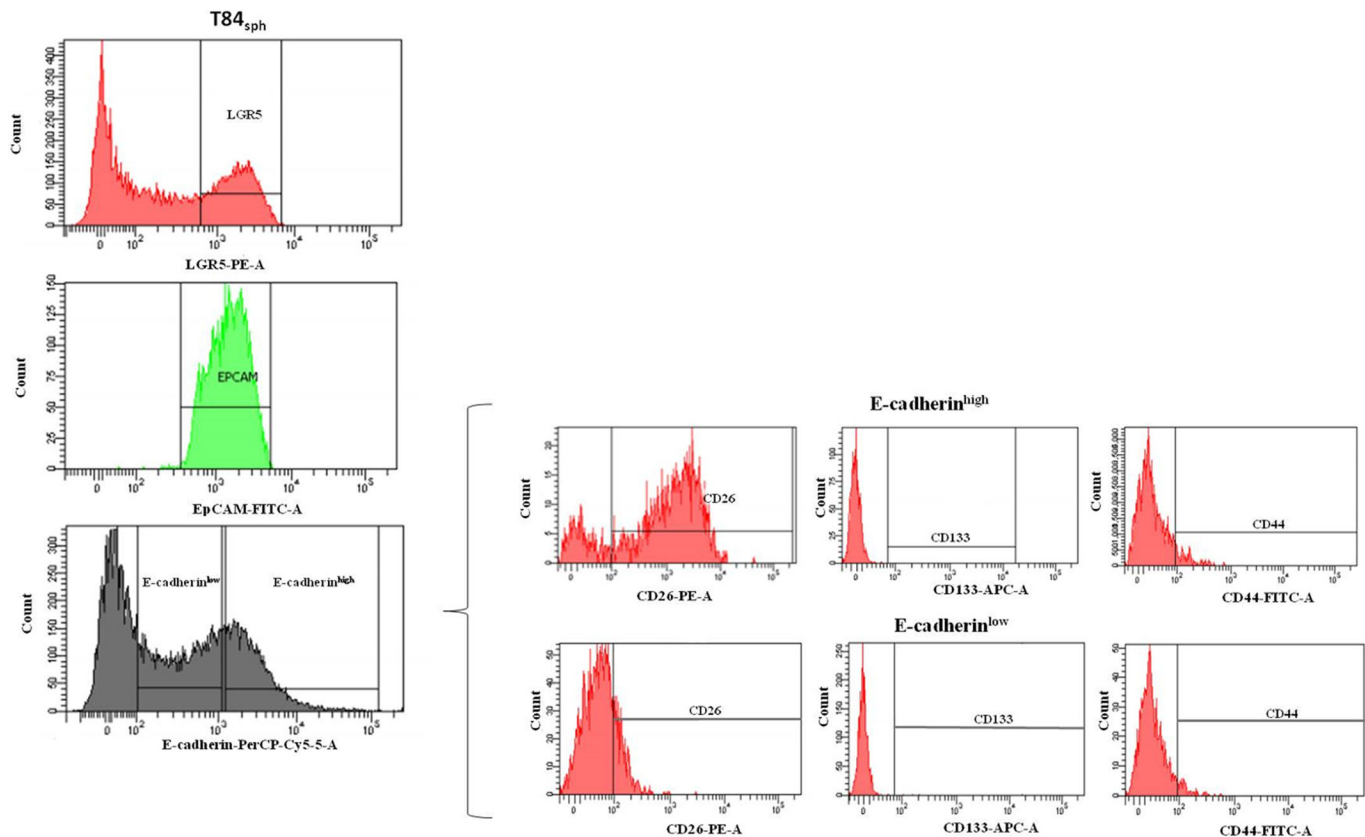
Several studies linked CD133<sup>high</sup> expression with a high risk of metastasis in CRC patients (LaBarge and Bissell, 2008; Kojima et al., 2008; Horst et al., 2008; Ong et al., 2010; Gallmeier et al., 2011), but the effective value of CD133 as a CSC biomarker is unclear, because, as observed in the SW620 colon cell line, sorted CD133<sup>+</sup> and CD133<sup>-</sup> subsets can undergo conversion between the two subsets (Hsu et al., 2013; LaBarge and Bissell, 2008; Kojima et al., 2008).

E-cadherin was used as a control of EMT. CSCs exist both in epithelial and mesenchymal states (Liu et al., 2014a,b), but EMT favors migration of cancer cells while inhibiting cell proliferation. Thus, MetSCs should be found in the epithelial state in the primary tumor, in the mesenchymal state in the peripheral blood (Oskarsson et al., 2014; Zimmerer et al., 2013), and in the epithelial state in the host organ. We assume that the loss of E-cadherin expression in sphere-derived cells (spheres were obtained over a 7-day period) is because they are newborn proliferating cells (that is, epithelial cells but lacking E-cadherin), rather than mesenchymal cells. However, the discrimination between sphere-derived small cells losing

E-cadherin expression and mesenchymal cells proliferating should be a priority of further research because these small cells, which completely lost EpCAM and LGR5 expression (recovered as they enlarged), were found in every passage of sphere-derived cell cultures, thus, data of LGR5 positivity could underestimate the frequency of intestinal CSCs in this work and *in vivo*. Interestingly, CD133, CD26 and CD44 expression remained in some of these small cells. The positivity for one or more of these markers was useful to identify each cell line, suggesting the presence of different lineages (e.g. E-cadherin<sup>-</sup>/CD133<sup>+</sup> was the origin of the CD133<sup>high</sup> subsets in the cell lines). In a very important study that described the stemness of spheroid-derived stem-like colon cancer cells from lines, the cells used in this work, these markers were not tested (Han et al., 2013).

EpCAM (CD326) overexpression is an early event during cancer progression in some types of tumors such as prostate and lung cancer, as well as in CRC (Miranda-Lorenzo et al., 2014; Pitule et al., 2014; Rowehl et al., 2014; Liu et al., 2014a,b). EpCAM appears in 85% of colorectal carcinomas, it can inhibit differentiation and promote proliferation (Hsu et al., 2013), and it is used to isolate CTCs in liquid biopsies (Dotse and Bian, 2016).

Although the presence of EpCAM<sup>high</sup>/CD44<sup>+</sup> cells correlated with the degree of differentiation, depth of invasion, clinical stage and metastatic status in CRC (Liu et al., 2014a,b) and gastric cancers



**Fig. 4. Sorting strategy of T84<sub>sph</sub>.** T84<sub>sph</sub> LGR5<sup>+</sup> (top histogram) and EpCAM<sup>high</sup> (middle histogram) were sorted into E-cadherin<sup>low</sup> and E-cadherin<sup>high</sup> subsets (bottom histogram). E-cadherin<sup>low</sup> and E-cadherin<sup>high</sup> subsets (on the right) were analyzed for CD26, CD133 and CD44 expression.

(Han et al., 2011), because of the plasticity and the not completely known role of their isoforms, CD44 expression does not seem to be an appropriate marker of MetSCs (Hsu et al., 2013; Pitule et al., 2014; Rowehl et al., 2014; Qiu et al., 2015; Nagano and Saya, 2004). However, in E-cadherin expressing cells, we observed that the frequency of the CD26<sup>+</sup>/CD44<sup>+</sup> subset increased independently of the expression of CD133.

LGR5 is a well-characterized marker of intestinal and colon stem cells (de Sousa e Melo et al., 2017; Kemper et al., 2012; Osawa et al., 2016; Zhang et al., 2016; Basu et al., 2016). We observed that sphere-derived cells showed a high frequency of the LGR5<sup>+</sup>/EpCAM<sup>high</sup> subset in many cell lines. This subset can represent clonogenic CSCs that proliferate originating the sphere (de Sousa e Melo et al., 2017; Zimmerer et al., 2013). However, in some cell lines, the frequency of LGR5<sup>+</sup>/EpCAM<sup>low</sup> cells was high and there were very low frequencies of EpCAM<sup>high</sup> found in SW480<sub>sph</sub>, SW620<sub>sph</sub> and COLO205<sub>sph</sub>. Interestingly, a small LGR5<sup>+</sup>/EpCAM<sup>-</sup> subset was detected in all lines<sub>sph</sub>, probably representing the CD26<sup>+</sup>/CD44<sup>+</sup>/CD66c<sup>+</sup> but EpCAM(CD326)<sup>-</sup> and CD133<sup>-</sup> CTC population that was an independent prognostic factor for CRC recurrence (Lieto et al., 2015).

Based on the fact that there are immune helper T cells with a defined CD26<sup>high</sup> phenotype (Krakauer et al., 2006), we tested if a similar subset with high staining of CD26 was present in CRC cell lines. Our data showed that the EpCAM<sup>high</sup>/LGR5<sup>+</sup> subset had high expression of E-cadherin and CD26, with variable frequency among the cell lines. The subset was mostly CD44<sup>+</sup>, and CD133<sup>+</sup> or CD133<sup>-</sup>, but there were also cells only CD26<sup>+</sup>. The frequency of this E-cadherin<sup>+</sup>/LGR5<sup>+</sup>/EpCAM<sup>high</sup>/CD26<sup>high</sup> lineage resembled that of the auto-fluorescent CSCs (Miranda-Lorenzo et al., 2014).

In addition, cell lines had cells only CD133<sup>+</sup>, particularly cell lines from early stages. The CD133<sup>high</sup>/CD26<sup>high</sup> and CD133<sup>high</sup>/CD26<sup>+</sup> subsets present in the cell lines (both CD44<sup>-</sup> and E-cadherin<sup>low</sup>, unlike the CD133<sup>+</sup>/CD26<sup>high</sup> subset) were not found in sphere-derived cells, suggesting additional clonogenic subsets that might be less related to the metastatic process (de Sousa e Melo et al., 2017).

In a previous study where the effect of chemotherapeutic adjuvants 5-fluorouracil, vinblastine, oxaliplatin, methotrexate, or irinotecan was studied on cell lines and in orthotopic tumor cells (Cutler et al., 2015), chemotherapy agents led to the loss of CXCR4<sup>+</sup>, another candidate marker in epithelial cancer cells (Miranda-Lorenzo et al., 2014; Jung et al., 2011; Rowehl et al., 2014; Cutler et al., 2015; Liao et al., 2010; Davies et al., 2015), in combination with CD133<sup>+</sup> (CXCR4<sup>+</sup>/CD133<sup>+</sup>), and to the enrichment of CD26<sup>+</sup>/CD44<sup>+</sup> cells, in agreement with our results, although our data show that these chemotherapy agents are not necessary to enhance CD26 and CD44 expression, as Cutler et al. suggest (Cutler et al., 2015). All in all, these data are supporting the role of CD26 in MetSCs (Cheung et al., 2017; Liao et al., 2010; Davies et al., 2015; Lam et al., 2014; Nishikawa et al., 2015; Collura et al., 2013; Gemei et al., 2013; Grunt et al., 2015) and the presence of at least two lineages of CSCs (Basu et al., 2016).

CD26 is related to some extent to the CXCR4/SDF-1 axis, because SDF-1 (CXCL-12) is a substrate of CD26/DPP4 enzymatic activity. However, most of the markers studied, including CD26, are related to cell-cell adhesion directly (as E-cadherin or EpCAM) or indirectly through the ECM. CD26 is known to associate with ADA, fibronectin and collagen (Cheng et al., 2003; Ghersi et al., 2002; Naim et al., 1999), and its binding to proteoglycans versican and

glypican-3 has been suggested (Havre et al., 2013; Khurana et al., 2013). CD44 is a receptor for hyaluronic acid (Liu et al., 2014a,b; Han et al., 2011). Nevertheless, whereas CD44 has functions only in cell adhesion and signaling, CD26 also has DPP4 activity in the intestinal brush border of normal mucosa with a role in protein digestion (Naim et al., 1999). To note, most primary tumors lose CD26 expression, and this may explain the difference between CD44 and CD26 staining of cell lines, although both can be cleaved from the cell surface (Nagano and Saya, 2004; Cordero et al., 2009). In fact, we demonstrated recently that soluble CD26 levels (sCD26) were a much better serum marker for the detection of CRC metastasis or tumor recurrence compared to other markers in clinical use, such as CEA, CA-19.9, or CA-72.4 (De Chiara et al., 2014). At the same time, a relationship between the presence of CD26<sup>+</sup> cells, detected by immunohistochemistry in primary CRC tumor biopsies, and the prognosis of metastasis has been demonstrated (Grunt et al., 2015), as well as between the presence of CD26<sup>+</sup>/CD44<sup>+</sup> and EpCAM<sup>-</sup>/CD133<sup>-</sup> CTC population, detected in liquid biopsy, and the prognosis of colorectal cancer recurrence (Lieto et al., 2015). It remains to be elucidated if these cells or the LGR5<sup>high</sup>/E-cadherin<sup>high</sup>/EpCAM<sup>high</sup>/CD26<sup>high</sup> cells described here cause metastasis in mice as the CD26<sup>+</sup>/CD133<sup>+</sup>/E-cadherin<sup>low</sup> CSC population (Pang et al., 2010).

CD26 and/or dipeptidyl peptidase 4 inhibitors can prevent colon cancer and lung metastasis in animal models (Angevin et al., 2017; Jang et al., 2015; Femia et al., 2013; Yorifuji et al., 2016), although epidemiological studies in humans are not clear (Enz et al., 2019). There are also encouraging results regarding anti-CD26 Ab in mesothelioma, renal and urological tumors (Enz et al., 2019). Overall, studies investigating CD26<sup>+</sup> subsets may ultimately contribute to the development of new treatment options for CRC (Cheung et al., 2017; LaBarge and Bissell, 2008; Han et al., 2013, 2011; Qiu et al., 2015; Osawa et al., 2016; Cutler et al., 2015; Khurana et al., 2013).

## MATERIALS AND METHODS

### Cell lines and culture conditions

Eight human colon cancer cell lines, SW1116, SW480, DLD-1, SW620, HT-29, Caco-2, COLO205 and T84 were obtained from the American Type Culture Collection (ATCC). SW1116, SW480, DLD-1, SW620, HT-29 and Caco-2 were cultured in DMEM media (Lonza) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich), 1% L-glutamine (Sigma-Aldrich) and 1% penicillin/streptomycin (Sigma-Aldrich). COLO205 was maintained in RPMI 1640 media (Lonza) supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine and 1% penicillin/streptomycin. T84 was maintained in DMEM/Ham's F12 media (Lonza) supplemented with 10% FBS, 1% L-glutamine and 1% penicillin/streptomycin. Cells were grown at 37°C in a humidified atmosphere of 5% CO<sub>2</sub>.

Dukes' stage and tissue origin for each cell line is shown in Table S1.

### Sphere formation assay

Sphere formation media was composed of serum-free DMEM/F12 (1:1), 20 ng/ml epidermal growth factor (EGF; Calbiochem), 10 ng/ml basic fibroblast growth factor (bFGF; Calbiochem), 5 µg/ml insulin (Sigma-Aldrich), 1x B-27<sup>®</sup> Supplement without Vitamin A (Gibco) and 1% penicillin/streptomycin (Sigma-Aldrich). For sphere suspension culture, cell lines grown in a two-dimensional monolayer were digested with trypsin (Lonza) and seeded at a density of 1×10<sup>4</sup> cells/ml in serum-free medium (SFM), in 100 mm ultra-low attachment plates (Corning) at 37°C and in a

humidified atmosphere of 5% CO<sub>2</sub>. The number and diameter of spheres were evaluated after 7 days. Medium was supplemented every alternate day to maintain proliferation and viability in all plates.

For serial passages, 7-day-old spheres were harvested and dissociated into single cells with trypsin. Dissociated cells were replaced in a new plate and cultured for 7 days. Efficiency of self-renewal was calculated from the number of cells formed in each passage from one single cell reseeded (100 cells/well).

### Antibodies

Antibodies used for western blot, immunofluorescence and flow cytometry were as follows: anti-CD26 (≠AF1180, R&D Systems), anti-CD26 (≠H00001803-D1 Novus Biologicals), anti-E-cadherin (≠610181, BD Biosciences), anti-vimentin (≠MA5-11883, Thermo Fisher Scientific Pierce), anti-EpCAM (≠2929, Cell Signaling Technology), anti-LGR5 (≠TA503316, OriGene Technologies), anti-CD133 (≠MAB4399, Millipore), anti-CD44-FITC (≠44F2, Immunostep), anti-CD133-APC (≠AC133 Miltenyi Biotech), anti-EpCAM-FITC (≠130-098-113, Miltenyi Biotech), anti-LGR5-PE (≠1030-100848, Miltenyi Biotech), anti-CD26-PE (≠26PE, Immunostep), anti-CD26-FITC (≠26F, Immunostep) and anti-E-cadherin-PerCP-Cy5.5 (≠563573, BD Biosciences).

Secondary antibodies used for western blot were horseradish peroxidase (HRP)-conjugated antibodies (Sigma-Aldrich). Secondary antibodies used for immunofluorescence were goat anti-rabbit AlexaFluor<sup>®</sup>488 and goat anti-mouse AlexaFluor<sup>®</sup>594 (Thermo Fisher Scientific).

### Western blot analysis

Cells were lysed in RIPA buffer (50 mM Tris-HCl, pH 7.5; 150 mM NaCl; 1% NP-40; 0.5% sodium deoxycholate and 0.1% SDS), supplemented with protease inhibitor cocktail (Roche). Insoluble components were removed by centrifugation and protein concentrations were measured using the Bradford protein assay (Bio-Rad). Equal amounts of protein (20 µg) were separated by SDS-PAGE and transferred onto PVDF membranes (Immobilon-P, Millipore). Membranes were blocked with 5% non-fat dry milk in PBST (0.1% Tween-20) for 1 h and probed with the primary antibodies diluted in PBST/5% non-fat dry milk. After washing, membranes were incubated with secondary horseradish peroxidase-conjugated antibodies. Protein signals were visualized with the Clarity Western ECL Substrate (Bio-Rad) according to the manufacturer's protocol. Images were acquired using a ChemiDoc XRS+ system (Bio-Rad). Quantitative image analysis was performed with Image lab software (Bio-Rad). Full-length blots of expression of different markers in the eight human colon cancer cell lines analyzed are showed in Fig S7.

### Immunofluorescence

Culture cells were grown in appropriate medium on glass coverslips until 70–80% confluence. Cells were fixed with 4% paraformaldehyde for 30 min at 37°C followed by permeabilization with 0.5% Triton X-100-phosphate buffered saline (PBS) for 3 min at room temperature. Cells were blocked with 2% bovine serum albumin in PBS for 30 min. Then cells were incubated with the appropriate primary antibodies for 1 h. After several washes, cells were incubated with appropriate fluorescent secondary antibodies for 45 min in the dark. Cell nuclei were stained with DAPI (1 µg/ml). The coverslips were mounted with MOWIOL on microscope slides and immunofluorescence was visualized using a fluorescence microscope (Olympus-BX51).



## Flow cytometry and cell sorting analysis

Human colon cancer cells derived from monolayer cultures and sphere-derived cells on day 7 after primary culture were adjusted to a final concentration of  $10^6$  cells/ml. Cell suspensions were incubated for 10 min with blocking solution and with appropriate antibodies in the dark at 4°C for 30 min. Cells were then washed with PBS and analyzed by flow cytometry (BD Accuri™ C6, BD Biosciences). We repeated each characterization three times to validate the results observed. We followed the protocol of Miranda-Lorenzo et al. (2014) to acquire autofluorescent cells from cultured cells: 30 μM riboflavin (Sigma-Aldrich) was added to cultured cells. For FACS acquisition cells were incubated overnight at 37°C, centrifuged at 300 g for 5 min and cell pellets re-suspended in PBS. Autofluorescent cells were excited with 488–561 nm laser and selected as the intersection with filters 496/578. Propidium iodide (Sigma-Aldrich) was used for exclusion of dead cells. Cell sorting was performed in a FACSAria Iiu analyzer (BD Biosciences) by using the PC FACSDiva software program (BD Biosciences).

## Gating strategies

Cells were gated on physical parameters (forward-scatter versus side scatter) to exclude dead or apoptotic cells, cell debris and aggregated cells. Single cells were gated on FSC-Area versus FSC-Height profile for excluding doublets.

An autofluorescence analysis was done with unstained cells and the background level was also determined for each fluorochrome. This helped us to evaluate the spillover of the different fluorochromes. Compensation controls were included for each fluorochrome.

Isotype controls were used to mark positive or negative staining. These marks were properly changed when any subset with high expressions (over the usual expression) of some CSC markers were observed in the different populations studied.

## Acknowledgements

Special thanks are given to Beatriz Sánchez Correa and Mercedes Peleteiro Olmedo for her technical assistance.

## Competing interests

The authors declare no competing or financial interests.

## Author contributions

Conceptualization: L.V.-I., M.P.d.I.C., O.J.C.; Methodology: L.V.-I., L.B.-C., M.R.-Q.; Formal analysis: M.R.-Q., M.P.d.I.C., O.J.C.; Investigation: L.V.-I., L.B.-C., M.R.-Q.; Resources: M.P.d.I.C.; Data curation: L.V.-I., M.R.-Q., M.P.d.I.C., O.J.C.; Writing - original draft: L.V.-I., O.J.C.; Writing - review & editing: L.V.-I., M.P.d.I.C., O.J.C.; Supervision: M.P.d.I.C., J.R.-B., O.J.C.; Funding acquisition: J.R.-B.

## Funding

This work was done with the Xunta de Galicia grants (supported by the: European Regional Development Fund (ERDF): Axudas consolidación e estruturación de unidades de investigación competitiva [GRC2014/019], Galician Network for Colorectal Cancer Research (REGICC) [R2014/039] and Agrupación estratégica InBiomed [2012/273].

## Data availability

Data sharing is not applicable to this article. Generated datasets are shown as tables. Original flow cytometry files analyzed during the current study are available from the corresponding author on reasonable request.

## Supplementary information

Supplementary information available online at <http://bio.biologists.org/lookup/doi/10.1242/bio.041673.supplemental>

## References

Acosta, M., Pereira, J. and Arroz, M. (2016). Screening of carcinoma metastasis by flow cytometry: a study of 238 cases. *Cytometry Part B* **90**, 289-294. doi:10.1002/cyto.b.21258

- Angevin, E., Isambert, N., Trillet-Lenoir, V., You, B., Alexandre, J., Zalzman, G., Vielh, P., Farace, F., Valleix, F., Podoll, T. et al. (2017). First-in-human phase 1 of YS110, a monoclonal antibody directed against CD26 in advanced CD26-expressing cancers. *Br. J. Cancer* **116**, 1126-1134. doi:10.1038/bjc.2017.62
- Basu, S., Haase, G. and Ben-Ze'ev, A. (2016). Wnt signaling in cancer stem cells and colon cancer metastasis. *F1000Res* **19**, 5. doi:10.12688/f1000research.7579.1
- Cheng, H.-C., Abdel-Ghany, M. and Pauli, B. U. (2003). A novel consensus motif in fibronectin mediates dipeptidyl peptidase IV adhesion and metastasis. *J. Biol. Chem* **278**, 24600-24607. doi:10.1074/jbc.M303424200
- Cheung, A. H., Iyer, D. N., Lam, C. S., Wong, S. K.M., Lee, H. S., Wan, T., Man, J., Chow, A. K. M., Poon, R. T., Pang, R. et al. (2017). Emergence of CD26+ cancer stem cells with metastatic properties in colorectal carcinogenesis. *Int. J. Mol. Sci.* **18**, E1106. doi:10.3390/ijms18061106
- Collura, A., Marisa, L., Trojan, D., Buhard, O., Lagrange, A., Saget, A., Bombled, M., Méchighel, P., Ayadi, M., Muleris, M. et al. (2013). Extensive characterization of sphere models established from colorectal cancer cell lines. *Cell Mol. Life Sci.* **70**, 729-742. doi:10.1007/s00018-012-1160-9
- Cordero, O. J., Salgado, F. J. and Nogueira, M. (2009). On the origin of serum CD26 and its altered concentration in cancer patients. *Cancer Immunol. Immunother* **58**, 1723-1747. doi:10.1007/s00262-009-0728-1
- Cutler, M. J., Lowthers, E. L., Richard, C. L., Hajducek, D. M., Spagnuolo, P. A. and Blay, J. (2015). Chemotherapeutic agents attenuate CXCL12-mediated migration of colon cancer cells by selecting for CXCR4-negative cells and increasing peptidase CD26. *BMC Cancer* **15**, 882. doi:10.1186/s12885-015-1702-2
- Davies, S., Beckenkamp, A. and Buffon, A. (2015). CD26 a cancer stem cell marker and therapeutic target. *Biomed. Pharmacother.* **71**, 135-138. doi:10.1016/j.biopha.2015.02.031
- De Chiara, L., Rodríguez-Piñero, A. M., Cordero, O. J., Vázquez-Tuñas, L., Ayude, D., Rodríguez-Berocal, F. J. and De la Cadena, M. P. (2014). Postoperative serum levels of sCD26 for surveillance in colorectal cancer patients. *PLoS ONE* **9**, e107470. doi:10.1371/journal.pone.0107470
- de Sousa e Melo, F., Kurtova, A. V., Harnoss, J. M., Kljavin, N., Hoeck, J. D., Hung, J., Anderson, J. E., Storm, E. E., Modrusan, Z., Koepfen, H. et al. (2017). A distinct role for Lgr5+ stem cells in primary and metastatic colon cancer. *Nature* **543**, 676-680. doi:10.1038/nature21713
- Dotse, E. and Bian, Y. (2016). Isolation of colorectal cancer stem-like cells. *Cytotechnology* **68**, 609-619. doi:10.1007/s10616-014-9806-0
- Driessens, G., Beck, B., Caauwe, A., Simons, B. D. and Blanpain, C. (2012). Defining the mode of tumour growth by clonal analysis. *Nature* **488**, 527-530. doi:10.1038/nature11344
- Enz, N., Vliegen, G., De Meester, I. and Jungraithmayr, W. (2019). CD26/DPP4 -a potential biomarker and target for cancer therapy. *Pharmacol. Ther.* **198**, 135-159. doi:10.1016/j.pharmthera.2019.02.015
- Femia, A. P., Raimondi, L., Maglieri, G., Lodovici, M., Mannucci, E. and Caderni, G. (2013). Long-term treatment with Sitagliptin, a dipeptidyl peptidase-4 inhibitor, reduces colon carcinogenesis and reactive oxygen species in 1,2-dimethylhydrazine-induced rats. *Int. J. Cancer* **133**, 2498-2503. doi:10.1002/ijc.28260
- Gallmeier, E., Hermann, P. C., Mueller, M.-T., Machado, J. G., Ziesch, A., De Toni, E. N., Palagyi, A., Eisen, C., Ellwart, J. W., Rivera, J. et al. (2011). Inhibition of ataxia telangiectasia- and Rad3-related function abrogates the in vitro and in vivo tumorigenicity of human colon cancer cells through depletion of the CD133(+) tumor-initiating cell fraction. *Stem Cells* **29**, 418-429. doi:10.1002/stem.595
- Gemei, M., Di, Noto, R., Mirabelli, P. L. and Del Vecchio, L. (2013). Cytometric profiling of CD133+ cells in human colon carcinoma cell lines identifies a common core phenotype and cell type-specific mosaics. *Int. J. Biol. Markers* **28**, 267-273. doi:10.5301/IJBM.5000020
- Gherzi, G., Dong, H., Goldstein, L. A., Yeh, Y., Hakkinen, L., Larjava, H. S. and Chen, W.-T. (2002). Regulation of fibroblast migration on collagenous matrix by a cell surface peptidase complex. *J. Bio. Chem.* **277**, 29231-29241. doi:10.1074/jbc.M202770200
- Grunt, T. W., Hebar, A., Laffer, S., Wagner, R., Peter, B., Herrmann, H., Graf, A., Bilban, M., Posch, M., Hoermann, G. et al. (2015). Prominin-1 (CD133, AC133) and dipeptidyl-peptidase IV (CD26) are indicators of infinite growth in colon cancer cells. *Am. J. Cancer Res.* **5**, 560-574.
- Han, M.-E., Jeon, T.-Y., Hwang, S.-H., Lee, Y.-S., Kim, H.-J., Shim, H.-E., Yoon, S., Baek, S.-Y., Kim, B.-S., Kang, C.-D. et al. (2011). Cancer spheres from gastric cancer patients provide an ideal model system for cancer stem cell research. *Cell. Mol. Life Sci.* **68**, 3589-3605. doi:10.1007/s00018-011-0672-z
- Han, X. Y., Wei, B., Fang, J. F., Zhang, S., Zhang, F. C., Zhang, H. B., Lan, T. Y., Lu, H. Q. and Wei, H. B. (2013). Epithelial-mesenchymal transition associates with maintenance of stemness in spheroid-derived stem-like colon cancer cells. *PLoS ONE* **8**, e73341. doi:10.1371/journal.pone.0073341
- Havre, P. A., Dang, L. H., Ohnuma, K., Iwata, S., Morimoto, C. and Dang, N. H. (2013). CD26 expression on T-anaplastic large cell lymphoma (ALCL) line Karpas 299 is associated with increased expression of versican and MT1-MMP and enhanced adhesion. *BMC Cancer* **13**, 517. doi:10.1186/1471-2407-13-517

- Horst, D., Kriegl, L., Engel, J., Kirchner, T. and Jung, A. (2008). CD133 expression is an independent prognostic marker for low survival in colorectal cancer. *Br. J. Cancer*. **99**, 1285-1289. doi:10.1038/sj.bjc.6604664
- Hsu, C.-S., Tung, C.-Y., Yang, C.-Y. and Lin, C.-H. (2013). Response to stress in early tumor colonization modulates switching of CD133-positive and CD133-negative subpopulations in a human metastatic colon cancer cell line, SW620. *PLoS ONE* **8**, e61133. doi:10.1371/journal.pone.0061133
- James, M. I., Howells, L. M., Karmokar, A., Higgins, J. A., Greaves, P., Cai, H., Dennison, A., Metcalfe, M., Garcea, G., Lloyd, D. M. et al. (2015). Characterization and propagation of tumor initiating cells derived from colorectal liver metastases: trials, tribulations and a cautionary note. *PLoS ONE* **10**, e0117776. doi:10.1371/journal.pone.0117776
- Jang, J.-H., Baerts, L., Waumans, Y., De Meester, I., Yamada, Y., Limani, P., Gil-Bazo, I., Weder, W. and Junggraithmayr, W. (2015). Suppression of lung metastases by the CD26/DPP4 inhibitor Vildagliptin in mice. *Clin. Exp. Metastasis*. **32**, 677-687. doi:10.1007/s10585-015-9736-z
- Jung, P., Sato, T., Merlos-Suárez, A., Barriga, F. M., Iglesias, M., Rossell, D., Auer, H., Gallardo, M., Blasco, M. A., Sancho, E. et al. (2011). Isolation and in vitro expansion of human colonic stem cells. *Nat. Med.* **17**, 1225-1227. doi:10.1038/nm.2470
- Kemper, K., Prasetyanti, P. R., De Lau, W., Rodermond, H., Clevers, H. and Medema, J. P. (2012). Monoclonal antibodies against Lgr5 identify human colorectal cancer stem cells. *Stem Cells* **30**, 2378-2386. doi:10.1002/stem.1233
- Khoja, L., Lorigan, P., Dive, C., Keilholz, U. and Fusi, A. (2015). Circulating tumor cells as tumor biomarkers in melanoma: detection methods and clinical relevance. *Ann. Oncol.* **26**, 33-39. doi:10.1093/annonc/mdl207
- Khurana, S., Margamuljana, L., Joseph, C., Schouteden, S., Buckley, S. M. and Verfaillie, C. M. (2013). Glypican-3-mediated inhibition of CD26 by TFPI: a novel mechanism in hematopoietic stem cell homing and maintenance. *Blood* **121**, 2587-2595. doi:10.1182/blood-2012-09-456715
- Kojima, M., Ishii, G., Atsumi, N., Fujii, S., Saito, N. and Ochiai, A. (2008). Immunohistochemical detection of CD133 expression in colorectal cancer: a clinicopathological study. *Cancer Sci*. **99**, 1578-1583. doi:10.1111/j.1349-7006.2008.00849.x
- Krakauer, M., Sorensen, P. S. and Sellebjerg, F. (2006). CD4 (+) memory T cells with high CD26 surface expression are enriched for Th1 markers and correlate with clinical severity of multiple sclerosis. *J. Neuroimmunol.* **181**, 157-164. doi:10.1016/j.jneuroim.2006.09.006
- LaBarge, M. A. and Bissell, M. J. (2008). Is CD133 a marker of metastatic colon cancer stem cells? *J. Clin. Invest.* **118**, 2021-2024. doi:10.1172/JCI36046
- Lam, C. S.-C., Cheung, A. H.-K., Wong, S. K.-M., Wan, T. M.-H., Ng, L., Chow, A. K., Cheng, N. S., Pak, R. C., Li, H. S. and Man, J. H. (2014). Prognostic significance of CD26 in patients with colorectal cancer. *PLoS ONE* **9**, e98582. doi:10.1371/journal.pone.0098582
- Liao, Y., Hu, X., Huang, X. and He, C. (2010). Quantitative analyses of CD133 expression facilitate researches on tumor stem cells. *Bio. Pharm. Bull.* **33**, 738-742. doi:10.1248/bpb.33.738
- Lieto, E., Galizia, G., Orditura, M., Romano, C., Zamboli, A., Castellano, P., Mabilia, A., Auricchio, A., De Vita, F. and Gemei, M. (2015). CD26-positive/CD326-negative circulating cancer cells as prognostic markers for colorectal cancer recurrence. *Oncol. Lett.* **9**, 542-550. doi:10.3892/ol.2014.2749
- Liu, D., Sun, J., Zhu, J., Zhou, H., Zhang, X. and Zhang, Y. (2014a). Expression and clinical significance of colorectal cancer stem cell marker EpCAM(high)/CD44(+) in colorectal cancer. *Oncol. Lett.* **7**, 1544-1548. doi:10.3892/ol.2014.1907
- Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., Martin-Trevino, R., Shang, L., McDermott, S. P. and Landis, M. D. (2014b). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell. Reports* **2**, 78-91. doi:10.1016/j.stemcr.2013.11.009
- Miranda-Lorenzo, I., Dorado, J., Lonardo, E., Alcalá, S., Serrano, A. G., Clausell-Tormos, J., Cioffi, M., Megias, D., Zagorac, S. and Balic, A. (2014). Intracellular autofluorescence: a biomarker for epithelial cancer stem cells. *Nat. Methods* **11**, 1161-1169. doi:10.1038/nmeth.3112
- Nagano, O. and Saya, H. (2004). Mechanism and biological significance of CD44 cleavage. *Cancer Sci*. **95**, 930-935. doi:10.1111/j.1349-7006.2004.tb03179.x
- Naim, H. Y., Joberty, G., Alfalah, M. and Jacob, R. (1999). Temporal association of the N- and O-linked glycosylation events and their implication in the polarized sorting of intestinal brush border sucrose-isomaltase, aminopeptidase N, and dipeptidyl peptidase IV. *J. Biol. Chem* **274**, 17961-17967. doi:10.1074/jbc.274.25.17961
- Nishikawa, S., Konno, M., Hamabe, A., Hasegawa, S., Kano, Y., Fukusumi, T., Satoh, T., Takiguchi, S., Mori, M., Doki, Y. et al. (2015). Surgically resected human tumors reveal the biological significance of the gastric cancer stem cell markers CD44 and CD26. *Oncol. Lett.* **9**, 2361-2367. doi:10.3892/ol.2015.3063
- Ong, C. W., Kim, L. G., Kong, H. H., Low, L. Y., Iacopetta, B., Soong, R. and Salto-Tellez, M. (2010). CD133 expression predicts for non-response to chemotherapy in colorectal cancer. *Mod. Pathol.* **23**, 450-457. doi:10.1038/modpathol.2009.181
- Osawa, H., Takahashi, H., Nishimura, J., Ohta, K., Haraguchi, N., Hata, T., Yamamoto, H., Mizushima, T., Takemasa, I., Doki, Y. et al. (2016). Full-length LGR5-positive cells have chemoresistant characteristics in colorectal cancer. *Br. J. Cancer*. **114**, 1251-1260. doi:10.1038/bjc.2016.112
- Oskarsson, T., Batlle, E. and Massagué, J. (2014). Metastatic stem cells: sources, niches, and vital pathways. *Cell Stem Cell* **14**, 306-321. doi:10.1016/j.stem.2014.02.002
- Pang, R., Law, W. L., Chu, A. C. Y., Poon, J. T., Lam, C. S. C., Chow, A. K. M., Ng, L., Cheung, L. W., Lan, X. R. and Lan, H. Y. (2010). A subpopulation of CD26<sup>+</sup> cancer stem cells with metastatic capacity in human colorectal cancer. *Cell Stem Cell* **6**, 603-615. doi:10.1016/j.stem.2010.04.001
- Pantel, K., Brakenhoff, R. H. and Brandt, B. (2008). Detection, clinical relevance and specific biological properties of disseminating tumour cells. *Nat. Rev. Cancer* **8**, 329-340. doi:10.1038/nrc2375
- Pitule, P., Cedikova, M., Daum, O., Vojtisek, J., Vycital, O., Hosek, P., Treska, V., Hes, O., Kralickova, M. and Liska, V. (2014). Immunohistochemical detection of cancer stem cell related markers CD44 and CD133 in metastatic colorectal cancer patients. *Biomed. Res. Int.* **2014**, 432139. doi:10.1155/2014/432139
- Qiu, H., Fang, X., Luo, Q. and Ouyang, G. (2015). Cancer stem cells: a potential target for cancer therapy. *Cell. Mol. Life Sci* **72**, 3411-3424. doi:10.1007/s00018-015-1920-4
- Ren, F., Sheng, W. Q. and Du, X. (2013). CD133 a cancer stem cells marker, is used in colorectal cancers. *World J. Gastroenterol.* **19**, 2603-2611. doi:10.3748/wjg.v19.i17.2603
- Rowehl, R. A., Burke, S., Bialkowska, A. B., Pettet, D. W., Rowehl, L., Li, E., Antoniou, E., Zhang, Y., Bergamaschi, R. and Shroyer, K. R. (2014). Establishment of highly tumorigenic human colorectal cancer cell line (CR4) with properties of putative cancer stem cells. *PLoS ONE* **9**, e99091. doi:10.1371/journal.pone.0099091
- Vanharanta, S. and Massagué, J. (2013). Origins of metastatic traits. *Cancer Cell*. **24**, 410-421. doi:10.1016/j.ccr.2013.09.007
- Varela-Calviño, R. and Cordero, O. J. (2015). Stem and immune cells in colorectal primary tumour: Number and function of subsets may diagnose metastasis. *World J. Immunol* **5**, 68-77. doi:10.5411/wji.v5.i2.68
- Wicha, M. S., Liu, S. and Dontu, G. (2006). Cancer stem cells: an old idea—a paradigm shift. *Cancer Res.* **66**, 1883-1890. doi:10.1158/0008-5472.CAN-05-3153
- Yorifuji, N., Inoue, T., Iguchi, M., Fujiwara, K., Kakimoto, K., Nouda, S., Okada, T., Kawakami, K., Abe, Y., Takeuchi, T. et al. (2016). The dipeptidyl peptidase-4 inhibitor sitagliptin suppresses mouse colon tumorigenesis in type 2 diabetic mice. *Oncol. Rep.* **35**, 676-682. doi:10.3892/or.2015.4429
- Zhang, S., Han, X., Wei, B., Fang, J. and Wei, H. (2016). RSP02 enriches LGR5(+) spheroid colon cancer stem cells and promotes its metastasis by epithelial-mesenchymal transition. *Am. J. Transl. Res* **8**, 354-364.
- Zimmerer, R. M., Korn, P., Demougin, P., Kampmann, A., Kokemüller, H., Eckardt, A. M., Gellrich, N. C. and Tavassol, F. (2013). Functional features of cancer stem cells in melanoma cell lines. *Cancer Cell Int.* **6**, 13-78. doi:10.1186/1475-2867-13-78